Processes Underlying Genetic Differentiation and Speciation in Orangutans (*Pongo* spp.)

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Alexander Nater

von

Zürich, ZH und Naters, VS

Promotionskomitee

Prof. Dr. Carel P. van Schaik (Vorsitz)

Dr. Michael Krützen (Leitung der Dissertation)

Zürich, 2012

Thesis Summary

A fundamental topic in evolutionary biology deals with the processes that led to genetic structuring of populations. Such a subdivision of gene pools allows local populations to follow independent evolutionary trajectories, during which the influence of mutation, genetic drift and natural selection might finally lead to speciation. The specific role of environmental processes in shaping the distribution of genetic diversity within and between species is still a topic of debate. The Southeast Asian Sundaland region is of special interest to study the influence of environmental factors on patterns of genetic diversity. This region consists of three major and numerous smaller islands, which have been recurrently connected to form a single landmass during Pleistocene glacial periods. The insular dynamics have led to an extremely high biodiversity on Sundaland, with a large number of endemic species on different islands.

In this thesis, I studied the processes influencing genetic structuring, gene flow and diversification in orangutans (*Pongo* spp.), the only Asian great apes, and species that are closely related to humans. Fossil evidence indicates that these apes were once widespread in southeastern Asia, but are now restricted to rainforests of northern Sumatra (*P. abelii*) and Borneo (*P. pygmaeus*). Furthermore, the fragmentation of their habitat renders them highly susceptible to extinction. Despite the remarkable features of orangutans, their evolutionary history is still very poorly understood.

For this project, I used the most extensive set of genetic samples from wild orangutans available to date to investigate the role of climate and sea level changes during Pleistocene glaciations, volcanic activity, geographic barriers and sex-biased dispersal behavior in the evolutionary history of orangutans. Using a combination of different genetic marker systems, I found remarkable genetic patterns in both orangutan species. Female-transmitted mitochondrial lineages exhibited strong geographic structuring, which I attributed to a combination of pronounced female philopatry and geographic dispersal barriers, such as rivers and mountain ridges. Remarkably, the Toba volcano on northern Sumatra led to a strong and long-lasting separation of mitochondrial gene pools in orangutans, as the only remaining Sumatran population south of the Toba caldera is more closely related to Bornean orangutans than to their Sumatran conspecifics.

In contrast to the strong geographic structure of mitochondrial haplotypes, Y-chromosomal markers, which are male-transmitted, revealed little geographic structure within each species and point toward a strong male bias in gene flow among orangutan populations. Male orangutans were able to distribute genes over large distances, as evidenced by my finding that regular genetic exchange between Bornean and Sumatran orangutans occurred until the beginning of the last glacial period ~110 kya.

Furthermore, by employing a novel modeling approach, I was able to shed light on the demographic history of the two orangutan species and investigate the impact of climate changes and anthropogenic factors on population sizes and connectedness. This approach revealed a marked difference in the demographic history of both orangutan species. Bornean orangutans were strongly influenced by recurrent changes in rainforest coverage during

Pleistocene glaciations, leading to at least one severe population bottleneck. In contrast, Sumatran orangutans displayed a remarkable stable demographic history until the end of the Pleistocene, when all Sumatran populations underwent a strong decline, most likely in response to increased hunting pressure by human colonizers.

In summary, I identified important environmental processes leading to genetic structuring and differentiation in orangutans. I showed that, under the influence of these environmental processes, orangutans experienced a complex demographic history, which led to distinct patterns of genetic variation. These findings will be of major importance for studies dealing with adaptive evolution in orangutans and other great apes.

Zusammenfassung der Dissertation

Eine fundamentale Fragestellung der evolutionären Biologie befasst sich mit den Prozessen, die zur genetischen Strukturierung von Populationen führen. Solche Unterteilungen des Genpools erlauben es lokalen Populationen unabhängige evolutionäre Entwicklungen zu durchlaufen, welche unter Einfluss von Mutationen, genetischer Drift und natürlicher Selektion schliesslich zu neuen Arten führen können. Die Rolle von Umweltprozessen auf die Strukturierung der genetischen Diversität innerhalb und zwischen Arten ist jedoch noch nicht im Detail erforscht. Die Sundaland-Region in Südostasien ist von besonderem Interesse um den Einfluss solcher Umweltfaktoren auf die Verteilung genetischer Diversistät zu untersuchen. Diese Region besteht aus drei grossen und zahlreichen kleineren Inseln, welche wiederholt während den glazialen Zyklen des Pleistozäns in einer einzigen grossen Landmasse verbunden waren. Diese Inseldynamik hat zu einer erstaunlichen Biodiversität in der Sundaland-Region geführt, die sich insbesondere durch eine hohe Anzahl an inselspezifischen endemischer Arten auszeichnet.

In dieser Dissertation habe ich die Prozesse untersucht, welche die genetische Strukturierung, den Genfluss und die Diversifizierung bei Orang-Utans (*Pongo* spp.) beeinflusst haben. Orang-Utans sind die einzigen asiatischen Menschenaffen und haben eine hohe phylogenetische Affinität zum Menschen. Fossilien belegen, dass diese Menschenaffen früher im südostasiatischen Raum weit verbreiten waren. Heutzutage kommen Orang-Utans allerdings nur noch in den Regenwäldern im Norden von Sumatra (*P. abelii*) und auf Borneo (*P. pygmaeus*) vor. Hinzu kommt, dass Orang-Utans durch die vom Menschen verursachte Fragmentierung ihres Habitats akut vom Aussterben bedroht sind. Trotz dieser interessanten Entwicklung ist die evolutionäre Geschichte der Orang-Utans noch weitgehend unbekannt.

Im Rahmen dieses Projektes habe ich die umfangreichste Sammlung genetischer Proben von wilden Orang-Utans verwendet, um die Rolle von pleistozänen Klimaveränderungen und Meeresspiegelschwankungen, vulkanischer Aktivität, geographischen Migrationsbarrieren und geschlechtsspezifischen Migrationsmustern in der evolutionären Geschichte der Orang-Utans zu untersuchen. Durch den Einsatz verschiedener genetischer Markersysteme konnte ich dabei bemerkenswerte Muster in beiden Orang-Utan-Arten finden. Die durch die Weibchen vererbte mitochondriale DNA zeigte eine ausgeprägte geographische

Strukturierung, welche ich mit einer Kombination von weiblicher Sesshaftigkeit und der Anwesenheit von Migrationsbarrieren wie Flüssen und Bergketten in Verbindung bringen konnte. Erstaunlicherweise führten die wiederholten Ausbrüche des Toba Vulkans im nördlichen Sumatra zu einer starken und langandauernden Unterteilung des mitochondrialen Genpools. Dies führte dazu, dass die einzige verbliebene Population südlich des Toba Kraters auf Sumatra auf mitochondrialer Ebene näher mit den Orang-Utans auf Borneo verwandt ist als mit den sumatranischen Populationen nördlich des Toba Kraters.

Im Unterschied zur mitochondrialen DNA zeigten die durch die Männchen vererbten Y-chromosomalen Marker sehr wenig geographische Struktur innerhalb der beiden Arten. Dies deutete klar auf ein ausgeprägt männchen-spezifisches Migrationsverhalten hin. Männliche Orang-Utans sind in der Lage Gene über lange Distanzen zu verbreiten, und meine Resultate deuten darauf hin, dass regelmässiger und von den Männchen verursachter Genfluss zwischen Borneo und Sumatra bis vor ca. 110‘000 Jahren stattgefunden hat.

Um die demografische Geschichte der beiden Orang-Utan-Arten sowie die Einflüsse von Klimaveränderungen und anthropogener Faktoren auf Populationsgrössen und Populationsstruktur detailliert zu untersuchen, verwendete ich neuartige Modellierungsverfahren. Dadurch könnte ich markante Unterschiede in der demografischen Geschichte der beiden Arten zeigen. Orang-Utans auf Borneo wurden stark beeinflusst durch die wiederkehrenden Veränderungen in der Ausbreitung des Regenwaldes während der glazialen Zyklen des Pleistozäns. Die Kontraktion des Regenwaldes während der Kaltzeiten zwang die Orang-Utans auf Borneo in ein Refugium und führte zu starken genetischen Flaschenhalseffekten. Im Kontrast dazu zeigten sumatranische Orang-Utans bis zum Ende des Pleistozäns eine stabile demografische Geschichte. Darauf folgte jedoch ein starker Einbruch in der Grösse aller Populationen auf Sumatra, welcher wahrscheinlich durch den verstärkten Jagddruck durch frühe menschliche Siedler ausgelöst wurde.

Zusammenfassend lässt sich sagen, dass im Rahmen dieser Dissertation erfolgreich wichtige Umwelteinflüsse identifiziert wurden, welche die genetische Strukturierung und Differenzierung innerhalb der Gattung der Orang-Utans beeinflusst haben. Ich konnte zeigen, dass Orang-Utans unter dem Einfluss dieser Prozesse eine komplexe demografische Geschichte durchlaufen haben, die zu prägnanten Mustern in der Verteilung der genetischen Variation geführt hat. Diese Resultate sind daher von grosser Bedeutung für zukünftige Studien über die adaptive Evolution von Orang-Utans und anderen Menschenaffen, inklusive der menschlichen Evolution.

Acknowledgments

I am very grateful to my supervisor Michael Krützen, who began with offering me an extremely exciting project with all the elements I was really keen on doing during my PhD. He did not only provide me with support whenever necessary but also left me enough room to shape this project according to my own wishes and interests.

I thank Carel van Schaik for his great interest in my work and his help in putting my results into the right context. His knowledge in many different fields of science is just amazing and provided valuable input at every stage of the project.

This PhD project would not have been the same without my office mates: Maja Greminger, Natasha Arora, Anna Kopps and Maria Jedensjö. I had a great time with you. Without our regular Badminton sessions, game nights and curry dinners, the life as a PhD student would have been much harder and certainly way more boring! Apart from all the not so scientific distractions, I really enjoyed our nerdy discussions in the office, which helped me enormously in getting a better idea of the meaning of all the data produced during the last four and a half years.

I enjoyed the great working atmosphere and spirit within the Evolutionary Genetics Group. A helping hand was always available and together we made some really exciting achievements. I will certainly remember the time as a member of the EGGs for a long time to come!

Working and living in Indonesia was certainly the most exciting but also most challenging part of my PhD project. Therefore, a special thank goes to Tony Weingrill, Ian Singleton, Serge Wich, Ellen Meulman, Andrea Permana, Fikar and Izumi for helping me to get along and succeed in the field. Thanks to their support, I never felt lost in the Indonesian culture, which is in many ways the complete opposite of what I was used to in Switzerland.

During the past years, I had the great opportunity to expand my knowledge in many fields of biology in fruitful discussions with various experts, including Daniel Wegmann, Lounès Chikhi, Lukas Keller, Homayoun Bagheri, Laurent Excoffier, Erik Meijaard, and Benoît Goossens. In one way or the other, these people all provided precious contributions to my PhD project. I am especially grateful to the external reviewers of this dissertation, Mark Beaumont and Michael E. Steiper, who provided valuable comments for improvement.

I thank all the members of the Anthropological Institute and Museum for their warm welcome when I started working here six years ago. The mixture of people from different fields of biology at the institute was always a great source of inspiration. I especially appreciated the help and support with administrative work by Claudia Zebib-Brunner.

Last but not least, I want to thank my family and friends for their great support during the four and a half years of my PhD. They always provided the right combination of motivation and soft pressure to conclude this project in a more or less reasonable amount of time. But most of all I want to thank my parents, who supported me in every possible way during my long time at university.

Table of Contents

List of Figures and Tables

Chapter 1: General Introduction

1.1. Why Reconstruct the Evolutionary History of Species?

Understanding the processes that have shaped the current distribution of genetic variation within and between species is one of the main goals of evolutionary biology. When a species no longer forms a single random mating unit but rather becomes genetically structured, mutation, drift and selection will act differently on the various local populations (Wright 1955). These processes may lead to genetic differentiation and ultimately speciation (Kimura & Weiss 1964). However, the extent to which drift and selection are acting upon local populations is heavily influenced by their demographic history. Genetic drift is a weak evolutionary force in large, panmictic populations, which allows selection to drive even slightly advantageous mutations to fixation. In contrast, in small or subdivided populations, random genetic drift is the dominating force which will determine the fate of newly arisen mutations largely independent of their selective value (Gillespie 2004). Thus, in order to reconstruct the evolutionary history of a species, we need to understand three main processes. First, we need to identify the factors that lead to the isolation of populations, which eventually allows them to follow independent evolutionary trajectories. Second, we need to reconstruct the demographic history of these distinct populations to be able to formulate expectations in terms of the extent to which different evolutionary forces shaped their genetic make-up. Third, we need to test specific genes that show marked geographic variation in allele frequencies within the range of a taxon for the presence of selective signals. This will allow us to determine the extent to which adaptations for specific local habitat conditions are driving the processes of speciation within a species.

This thesis aims to investigate the first two processes described above, that is the forces leading to a partitioning of a species' gene pool and the influence of demographic history on the evolutionary trajectory of local subpopulations, using orangutans as a study system. Another PhD project by Maja Greminger currently running at the Anthropological Institute and Museum of the University of Zurich is investigating the impact of selective pressures on the gene pools of different orangutan populations. Combined, both studies will help us to understand the processes that formed current patterns of genetic variation within and between two closely related species, which might eventually shed some light on the origin of the astonishing diversity of life forms on earth.

1.2. Orangutans in Sundaland as Study Taxon

In order to interpret the patterns of genetic variation that can be observed in a taxon, we need to have detailed knowledge about the habitat conditions and the geographic variation thereof over an evolutionarily relevant timeframe. Furthermore, we require a good understanding of the behavioral ecology, physiology and morphology of the species of interest. The choice of the study organisms is therefore crucial to be able to draw meaningful conclusions about the processes of diversification within and between species.

1.2.1. The Sunda Region as Model to Study Processes of Diversification

The Southeast Asian region of Sundaland includes the Malay Peninsula, the major islands of Sumatra, Borneo and Java, numerous smaller islands and the continental shelf that connects all the islands with the mainland (Molengraaff 1921). This region harbors an exceptionally high biodiversity, with around 15,000 endemic plant species and over 700 endemic vertebrate species (Myers et al. 2000; Sodhi et al. 2004). A multitude of processes are thought to have shaped these unique biogeographic patterns. During the Tertiary (65–1.8 Ma), tectonic plate movements caused frequent landmass reconfigurations, allowing plants and animals to colonize the Sunda islands from the Southeast Asian mainland during certain time spans, but also isolated parts of Sundaland for prolonged periods (Morley 2000; Meijaard 2004). The Sunda islands reached their current shape only in the Early Pleistocene (2.6–1.8 Ma) (Meijaard 2004). Starting in the Pleistocene, recurrent glacial and interglacial periods led to drastic sea level changes (Lisiecki & Raymo 2005). During glacial periods, and thus low sea levels, most of the continental shelf was exposed, opening up land bridges among the islands and to the mainland (Voris 2000, Figure 1.1). The drop in the mean annual temperatures during the glacial periods had, however, also a strong impact on the evergreen rainforest in the Sunda region (Morley 2000). The generally drier and more seasonal conditions led to a severe reduction of rainforest coverage, which might have become restricted to small refugia in proximity to the coast or along mountain slopes (Flenley 1998; Gathorne-Hardy et al. 2002; Cannon et al. 2009).

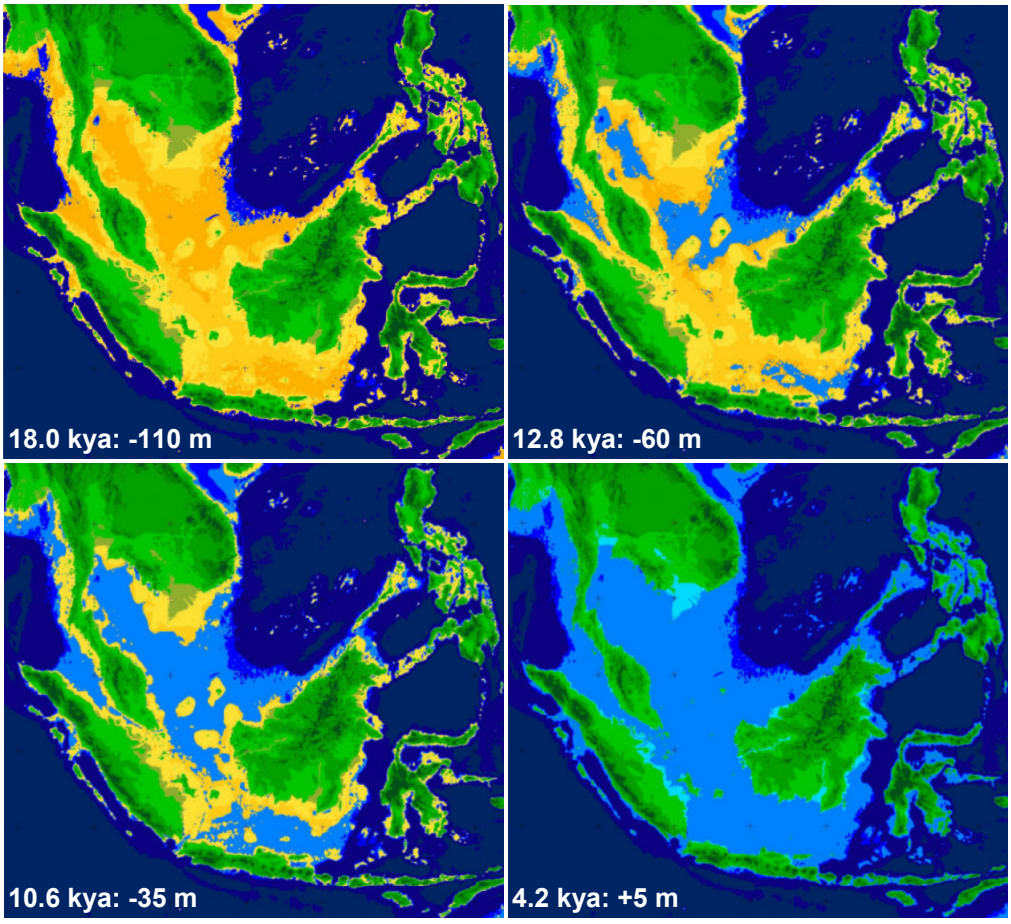


Figure 1.1: Effects of sea level changes on the exposed landmass in Sundaland since the last glacial maximum (maps adapted from Sathiamurthy & Voris 2006).

In addition to the impact of geological and climatic changes, the Sunda region has a rich volcanic history, as it is surrounded by subduction zones of tectonic plates (Hall 1996). Major volcanic eruptions might have shaped extant biogeographic patterns by causing local extinctions and subsequent recolonization. Of special notice here is the Toba caldera on northern Sumatra, which has seen at least four major eruptions during the last 1.2 million years (Chesner et al. 1991). The volcanic activity of the Toba volcano culminated in a supereruption around 73 kya, which was one of the most powerful explosive eruptions known in geological history and might have had a severe global impact on the climate (Rose & Chesner 1987; Chesner et al. 1991; Rampino & Ambrose 2000; Williams et al. 2009).

The Sunda region features a large diversity of habitats and climatic conditions, promoting its remarkable biodiversity. The climatic conditions on Sundaland are mainly influenced by the proximity to the equator, which leads to high annual rainfall and a prevailing vegetation of evergreen rainforest (Whitmore & Burnham 1984; Morley 2000). Rainfall is mostly associated with the monsoon winds coming from the Indian Ocean in northern summer and from the Pacific Ocean in northern winter (Harrison 2001). These winds take up moisture from the warm seas, raining down when the air rises up near the equator (Wang et al. 1999). Despite the high annual rainfall throughout Sundaland, seasonal and regional patterns of low rainfall occur, mostly in between the two monsoon phases, in regions where monsoon winds bring cold and dry continental air, or in regions that are located on the lee side of large mountain ranges (Wang et al. 1999; Whitten et al. 2000). These regional differences in seasonality have produced a high diversity of forest habitats, from near aseasonal rainforest in northwestern Sumatra to forests with pronounced wet and dry seasons in northeastern Borneo (MacKinnon et al. 1996; Whitten et al. 2000).

On a less predictable temporal scale, the El Niño Southern Oscillation (ENSO) phenomenon influences flora and fauna of the Sunda region profoundly (Rasmusson & Wallace 1983; Harrison 2001). This non-cyclical change in oceanic currents leads to a sharp drop in sea temperature in the western Pacific, which in turn reduces evaporation and rainfall in the eastern Sunda region, causing severe droughts and forest fires (Harrison 2001). El Niño events occur very irregularly in intervals of 2–7 years and the strength of the effects differs considerably among El Niño events (Rasmusson & Wallace 1983; Guilyardi et al. 2009). In addition to its impact on rainfall patterns, ENSO is associated with mast flowering in Southeast Asian dipterocarp forests (Ashton et al. 1988; Wich & Van Schaik 2000), which leads to community-wide peaks in seed production at the end of El Niño droughts, followed by long periods of extremely low productivity. This correlation between ENSO and masting is much stronger on the eastern part of the Sunda region as compared to northwestern Sumatra and western peninsular Malaysia (Ashton et al. 1988; Yasuda et al. 1999; Wich & Van Schaik 2000).

1.2.2. Orangutans as a Study Taxon for Genetic Diversification

Given the well-documented occurrence of such diverse forces, the Sundaland region is a perfect model system for studying the evolutionary processes that lead to speciation and endemism. Among the mammals endemic to Sundaland, orangutans (*Pongo* spp.), the only Asian great apes, seem to be well suited to investigate the processes leading to local differentiation and adaptation due to multiple reasons. First, as evidenced by the dramatic

changes in their distribution in Southeast Asia (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000), the evolutionary history of orangutans seems to have been strongly influenced by the aforementioned environmental processes.

Second, there are two recognized orangutan species, the Sumatran orangutan (*Pongo abelii*) and the Bornean orangutan (*Pongo pygmaeus*) (Groves 2001), that are currently geographically isolated but were historically linked during glacial periods (Voris 2000). These two species are genetically highly differentiated (Zhi et al. 1996; Kaessmann et al. 2001; Zhang et al. 2001; Locke et al. 2011), but have not yet reached reproductive isolation (Muir et al. 1998), thus allowing for gene flow at low sea levels. Studying genetic variation in closely related allopatric, that is geographically separated, species might reveal signals of adaptations that can be linked to local habitat conditions.

Third, the current range of orangutans on both islands is subdivided by rivers and mountain ridges, which act as dispersal barriers, promoting local genetic differentiation (Warren et al. 2001; Goossens et al. 2005; Arora et al. 2010). Within the patchy distribution of orangutans, a high diversity of habitat conditions is found, ranging from high-altitude dryland forests to coastal peat-swamp forests. Both factors, high intra-species genetic differentiation and high diversity of local habitats, might promote local adaptations and eventually further speciation in the genus *Pongo*.

Fourth, orangutans have an extremely slow life history (Wich et al. 2009a). The long generation time preserves the genetic signals of demographic changes for longer as compared to a fast reproducing species. Last, both Bornean and Sumatran orangutans have been extensively studied in the wild for over 40 years (MacKinnon 1974). Thus, we can capitalize on a profound knowledge of the behavioral ecology, morphology and physiology, as well as the geographical variation thereof in both orangutan species (Wich et al. 2009b). Furthermore, the habitat diversity over large parts of the current range of orangutans has been well characterized due to the presence of over a dozen long-term study sites on Borneo and Sumatra, allowing us to link certain traits to specific local conditions.

1.2.3. Distribution and Population History of Orangutans

Nowadays, wild orangutans exclusively inhabit rainforests on the islands of Borneo and Sumatra (MacKinnon 1974; Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). While orangutans have a rather wide-spread distribution on Borneo, they are restricted to small forest areas on the northern tip of Sumatra (Figure 1.2). However, the current distribution of orangutans is in no way representative of their past evolutionary history. The rich subfossil record provide evidence that orangutans were widespread in Southeast Asia during the Pleistocene, including populations on Java, in southern China, and probably even in northern India (Kahlke 1972; von Koenigswald 1982b; Rijksen & Meijaard 1999; Delgado & Van Schaik 2000).

Ecological and anthropogenic explanations have been invoked for the massive decline in the orangutan distribution (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). The distribution of orangutans has moved southwards in response to the shift of subtropical and tropical zones and the increase in seasonality during the Pleistocene (Jablonski 1998). Thus, ecological factors can explain the absence of orangutan in southern China, but not the

extinctions in the tropical zones of Indochina, Java and southern Sumatra. In fact, the expansion of tropical rainforest since the last glacial maximum (LGM, 19–26 kya) should have favored the persistence of orangutan populations in these areas (Morley 2000; Gathorne-Hardy et al. 2002; Bird et al. 2005). Extensive hunting by early human colonizers might be the main reason for the current patchy distribution of orangutans on the Sunda archipelago (Delgado & Van Schaik 2000). If so, the current distribution of orangutans was likely determined by the presence of refugia that were hard to reach for human hunters, such as swamp and mountain forests. Furthermore, biogeographic studies showed that hunting by early humans had a worse effect on the fauna of small islands as compared to larger islands or on the mainland (Steadman 1993; Reis & Garong 2001), which explains why orangutans are still present on Borneo and Sumatra, but not on any of the smaller islands of the Sunda archipelago. Beginning in the last century, large-scale deforestation by humans has led to a further decline and fragmentation of the remaining orangutan populations on both Borneo and Sumatra (Rijksen & Meijaard 1999; Wich et al. 2008).

Figure 1.2: Current distribution of orangutans in Southeast Asia (shaded areas) and Pleistocene subfossil sites (black dots) (map from Delgado & Van Schaik 2000).

The available genetic data point toward a complex history of colonization of the Sunda islands by orangutans. Sumatran orangutans show a higher genetic diversity than Bornean orangutans (Muir et al. 2000; Steiper 2006; Locke et al. 2011), even though their census size is about ten times smaller as compared to the Bornean species (Wich et al. 2008). It has thus been suggested that orangutans colonized the islands of Sundaland by entering Sumatra from the Southeast Asian mainland and spread via southern Sumatra to Java and Borneo (Rijksen & Meijaard 1999). The immigration routes would have been mostly defined by the large river systems dissecting the exposed Sunda shelf (Harrison et al. 2006, Figure 1.3). The Bornean populations were subsequently isolated from the remaining range of orangutans by the rising

sea levels during the Pliocene (5.3–2.6 Ma, Woodruff 2003), but conflicting evidence has been presented about the divergence time of the two orangutan species (Muir et al. 2000; Verschoor et al. 2004; Steiper 2006). The glacial periods of the Pleistocene (2.6–0.012 Ma) led to a fall in sea levels due to increased glaciation, cyclically exposing the Sunda shelf and opening up land bridges among the islands of the archipelago (Voris 2000; Lisiecki & Raymo 2005). This would have theoretically allowed orangutans to migrate between the islands, but the exposed shelf might not have been covered with suitable rainforest habitat, as glacial periods were also generally drier and more seasonal and the soils on large parts of the exposed shelf were most likely very nutrient-poor (Flenley 1998; Morley 2000; Gathorne-Hardy et al. 2002; Bird et al. 2005; Slik et al. 2011). The last opportunity for orangutans to cross the Sunda shelf would have been at the end of the last glacial period around 10 kya (Voris 2000), but evidence for such recent (natural) inter-island gene flow is ambiguous (Muir et al. 2000; Verschoor et al. 2004; Kanthaswamy et al. 2006; Steiper 2006).

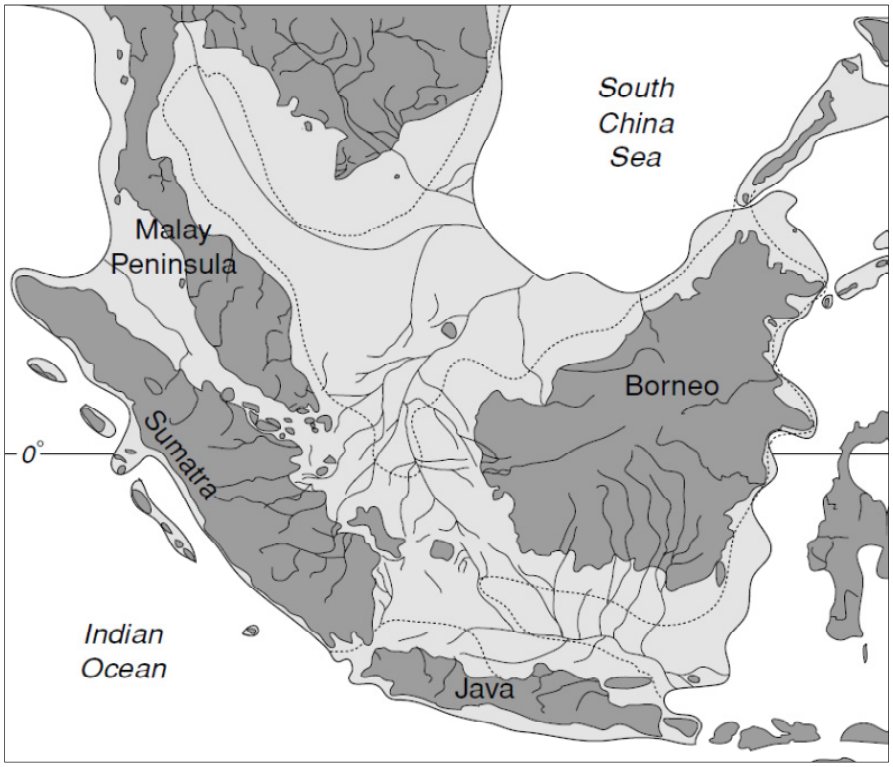


Figure 1.3: Reconstruction of paleo river systems on the exposed Sunda shelf during glacial sea level lowstands. The solid and dotted lines represent the extent of the exposed shelf at 100 m and 50 m below current sea level, respectively (map from Harrison et al. 2006).

Genetic evidence indicates the presence of demographic fluctuations, such as population contractions and expansions, for both Bornean and Sumatran orangutans during the Pleistocene (Steiper 2006; Locke et al. 2011). Such demographic changes in orangutan numbers could have different reasons. First, the reduced annual mean temperature and rainfall during the Pleistocene glacial periods caused a contraction of evergreen rainforest and therefore of suitable orangutan habitat, which might have pushed orangutans into mountain refugia (Flenley 1998; Morley 2000; Gathorne-Hardy et al. 2002). Second, major volcanic eruptions, mainly on Sumatra and Java, might have led to the extinction of local populations followed by recolonization from other populations (Chesner et al. 1991; Muir et al. 2000). In

particular the Toba eruption 73 kya on Northern Sumatra, which is considered to be the most powerful explosive volcanic eruption within the last 25 million years (Louys 2007), must have had severe consequences for the flora and fauna in the region and it is conceivable that it has decimated orangutans to small refugial populations. Lastly, prehistoric hunting by hunter-gatherer societies might have led to large-scale population collapses (Delgado & Van Schaik 2000).

1.2.4. Behavioral Ecology of Orangutans

Orangutans are extreme arboreal specialists, with adaptations for arboreality that are unique among great apes (Delgado & Van Schaik 2000). The diet of orangutans is predominantly frugivorous (Delgado & Van Schaik 2000; Morrogh-Bernard et al. 2009). This represents a major nutritional challenge, as fruit availability is subject to strong temporal variation in many parts of their distribution. To cope with periods of food scarcity, orangutans evolved manifold strategies, such as reliance on low-quality fallback foods like bark and leaves, an extremely low energy expenditure and the ability to store large amounts of fat (Knott 1998; Morrogh-Bernard et al. 2009; Pontzer et al. 2010). Concurrent with their low metabolic rate, orangutans exhibit an exceptionally slow life history, with a estimated generation time of 25 years and an interbirth interval ranging from 6–9 years (reviewed in Wich et al. 2009a).

The social structure among orangutans is best described as semi-solitary or non-gregarious (Delgado & Van Schaik 2000). Female orangutans tend to settle near their maternal kin with highly overlapping home ranges, leading to loose associations among females (Singleton & van Schaik 2002; van Noordwijk et al. 2012). In contrast, male orangutans have much larger home ranges (Singleton & van Schaik 2001), but tend to form consortships with females during which both individuals range in a coordinated way (Delgado & Van Schaik 2000). Behavioral and genetic studies indicate that orangutans have a strong sex bias in dispersal, with females showing strong philopatric tendencies, whereas males leave the natal area (Galdikas 1995; Singleton & van Schaik 2002; Morrogh-Bernard et al. 2011; Arora et al. 2012; Nietlisbach et al. 2012; van Noordwijk et al. 2012).

Unique among great apes, orangutans occur in two distinct sexually mature male morphs. The flanged males differ in morphology and behavior from their unflanged conspecifics (Delgado & Van Schaik 2000). Flanged males are much bigger and heavier than both unflanged adult males and females and exhibit secondary sexual characters such as cheek pads and a throat sac (Kingsley 1982). These sexual characters are associated with the generation of characteristic vocalizations, the so-called long calls (MacKinnon 1974). While these two morphs have first been described as subadult and adult stages of male development (MacKinnon 1974), more recent research point towards two alternative reproductive strategies, where unflanged males delay the development of secondary sexual characters depending on the dominance of other local males (Utami et al. 2002; van Schaik 2004; Utami Atmoko et al. 2009a). Unflanged males can use their better mobility compared to large and heavy flanged males to avoid encounters with them and regularly enforce copulation on females (MacKinnon 1974; Rijksen 1978). Genetic studies on wild orangutans have shown that both morphs sire offspring in both species (Utami et al. 2002; Goossens et al. 2006b). However, there are large differences between the two orangutan species regarding the

densities of flanged males and the number of fights between them (Delgado & Van Schaik 2000).

Orangutans are large-brained animals that show complex behavioral repertoires (van Schaik et al. 2003; van Schaik et al. 2009a), including tool use (van Schaik & Knott 2001). Many behavioral traits show distinct geographic variation, a pattern that has been described as cultural, whereas certain behaviors are locally innovated by an individual and then spread within the population and among populations via social learning (van Schaik et al. 2003; van Schaik et al. 2009a; Krützen et al. 2011). Thus, cultural evolution might have played a major role in the evolution of behavioral adaptations to the specific local habitat conditions found within the distribution of orangutans (reviewed in Wich et al. 2009b).

1.2.5. Conservation Aspects

Orangutans are among the most endangered great apes (IUCN 2012), with an estimated 54,000 Bornean and 6,600 Sumatran orangutans left in the wild (Wich et al. 2008). Orangutan population sizes of both species have decreased by a factor of at least ten since the beginning of the last century. This dramatic collapse has been mostly attributed to human activities, such as hunting, illegal pet trade and habitat loss due to deforestation and forest conversion (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000; van Schaik et al. 2001; Goossens et al. 2006a). Orangutans are especially prone to such disturbances due to their extremely slow life history (Wich et al. 2009a) and their strong dependency on intact rainforest habitat (Delgado & Van Schaik 2000). In addition to the reduction in population sizes, the human-induced habitation degradation led to a severe fragmentation of the habitat, often resulting in habitat patches that harbor only a few hundred orangutans (Wich et al. 2008).

Such small and isolated populations are highly vulnerable to future extinctions due to multiple reasons. First, the small number of individuals and the restricted dispersal possibilities increase the risk of mating among relatives, thus potentially leading to inbreeding depression (Hedrick & Kalinowski 2000). Second, small populations are more affected by genetic drift, which lowers the genetic diversity in the population, thereby reducing the potential to adapt to changes in the environment, such as the effects of global warming (Franklin 1980; Selander 1983). Third, genetic drift rather than selection is the dominating force determining the fate of mutations in small populations. Therefore, deleterious mutations might come to fixation instead of being eliminated by selection, thus reducing the average fitness of populations (Hedrick & Kalinowski 2000; Reed & Frankham 2003). Last, small populations are prone to be erased by diseases or natural disasters, such as the large-scale forest fires associated with El Niño droughts (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). Such detrimental effects on local populations in combination with the ongoing habitat degradation will undoubtedly lead to a further reduction in the number of orangutans (Wich et al. 2008). The grim outlook for the genus *Pongo* highlights the need for genetic methods to identify hotspots of genetic diversity, hidden population substructure and dispersal corridors to support efficient conservation actions.

1.3. Methods to Reconstruct Evolutionary History

Attempts to reconstruct the evolutionary history of a taxon usually involve multiple steps. The first goal is to describe how genetic diversity is geographically distributed within the extant range of the taxon of interest. Such analyses are often described with the term phylogeography (Avise 2000). Genetic markers are analyzed from geographically defined samples and used to reconstruct the genealogical histories of the analyzed genetic regions, the so-called gene trees. In a second step, detailed information about landscape features, geological history and the behavioral ecology of the taxon is used to investigate the processes that might have shaped the observed phylogeographic patterns.

1.3.1. Analytical Approaches

To identify specific processes that have shaped the phylogeographic patterns in a taxon, two distinct methods are commonly applied: molecular dating of coalescent events and comparing the shape of gene trees against the expectations under certain demographic scenarios (Knowles & Maddison 2002). In the first method, the estimates of coalescence times of specific lineages are used to link population splits with contemporary geological events, which might yield information about the processes leading to these population divisions. Building up a causal relationship between the coalescence of geographically structured lineages and population history is, however, problematic. The coalescence of two geographically delimited lineages in a gene tree does not correspond to the split time of an ancestral panmictic population into two sister populations (Maddison 1997; Nichols 2001; Knowles & Maddison 2002; Nielsen & Beaumont 2009). Rather, the relationship between coalescent events and population splitting times is determined by the effective population sizes and the migration rate between the two sister populations. Furthermore, the sequence of coalescent events of lineages might differ from the real order of population splits, especially if such splits happened in quick succession. This is caused by the fact that lineages might need a considerable amount of time to coalesce in the ancient population. Since the coalescent process is stochastic, older population splits might in fact produce more recent coalescence times than younger ones (Maddison 1997; Nielsen & Beaumont 2009). This problem can be alleviated by analyzing multiple independent genetic markers, which partly allows accounting for the stochasticity of the coalescent process. However, different independent marker systems will often yield inconsistent branching patterns, especially for comparatively recent population splits. In such cases, gene trees are often not easily interpretable and a sound statistical framework is needed to draw conclusions about the processes that shaped these gene trees (Knowles & Maddison 2002; Nielsen & Beaumont 2009).

Even if population split times can be accurately estimated, this information alone is not sufficient to accurately describe the population history of a taxon. Effective population size N_e and its change during time is of special importance for all questions dealing with the trajectories of specific genetic variants within populations. Genetic drift and selection both influence changes in allele frequencies, but the extent to which one force dominates over the other is strongly influenced by N_e (Gillespie 2004). Population bottlenecks in combination with local isolation can dramatically change the gene pool of a population and might be one of the most important factors in speciation (Mayr 1970). Such changes in population size and connectivity are, for example, often associated with local refugia during the Pleistocene

glaciations (e.g. Hewitt 1996). Summarizing the gene trees with test statistics and comparing these statistics against the expected values obtained under specific demographic scenarios might reveal signals of population size changes and allow testing such signals in a statistical framework (Knowles & Maddison 2002). The expected values for the test statistic under a null hypothesis are derived either analytically or by simulations, which explicitly takes the stochasticity of the coalescent process into account. Thus, by obtaining the distribution of values expected under a certain demographic scenario, threshold values can be deduced which allow rejecting a null hypothesis with a given level of statistical error. Such simple test statistics, however, often fail to yield satisfactory results (Knowles & Maddison 2002). First, the demographic event might leave only minor traces in the genealogies, which might, depending on the test statistic, not be distinguishable from the stochastic noise of the coalescent process. Second, these statistics summarize complete genealogical trees up to the most recent common ancestor (MRCA) of all samples. Since gene trees might have been affected by multiple demographic processes since the MRCA, these signals will all be overlaid in the test statistic and might cancel each other out.

A promising solution to the aforementioned problems of demographic reconstruction lies in model-based inference of phylogeography (Marjoram & Tavare 2006; Nielsen & Beaumont 2009; Beaumont et al. 2010). Model-based inference aims at finding a sufficient model and its underlying parameters that are able to produce the observed genetic data. Such models can be compared to alternative models in a statistical framework and the uncertainty of model choice and parameter estimates can be assessed (Beaumont et al. 2010). The parameter values of the model can be estimated by maximizing a likelihood function, which specifies the probability to obtain the observed data given the model and its parameter values. Therefore, model-based inference allows investigating demographic processes in a likelihood-based framework. However, this approach is not without problems. The tested demographic models need to be highly simplified in order to keep the number of estimated model parameters within reasonable limits and computational time manageable (Beaumont et al. 2002; Marjoram et al. 2003; Wegmann et al. 2009b). Assessing the impact of ignored aspects of real population history, such as population substructure and unsampled populations is not easy. For more realistic demographic models, the likelihood function can often not be calculated analytically (Marjoram et al. 2003).

A powerful solution to this problem is offered by Approximate Bayesian Computation (ABC), which aims to approximate the likelihood functions of complex demographic models rather than to solve them analytically (Tavare et al. 1997; Pritchard et al. 1999; Beaumont et al. 2002). In an ABC framework, the approximation of the likelihood function is accomplished by simulating a large number of data sets under a given model. The model parameters used for the simulations are drawn from prior distributions, which represent a priori knowledge about the demographic history. Summary statistics of the simulated data sets are then compared with the real observed data and simulations within a certain Euclidian distance to the real data are retained to estimate the posterior distributions of the model parameters. Therefore, this approach completely avoids the problems associated with the calculation of likelihood functions in complex demographic models. If the rejection distance is small enough, the densities of retained parameter values closely approximate the real posterior distributions. This allows building models that incorporate multiple populations with different

split times, rates of gene flow, effective population sizes and demographic changes in population size (e.g. Cornuet et al. 2010; Wegmann et al. 2010). Many different genetic marker systems can be analyzed together and combined into a common parameter estimation procedure. Furthermore, alternative models can be directly compared irrespective of the prior distributions of their model parameters and the certainty of model choice can be calculated in the form of Bayes factors (Kass & Raftery 1995).

ABC methods have been successfully applied to reconstruct demographic histories of different species, including great apes. For example, Fagundes et al. (2007) could evaluate the likelihood of different models of human evolution using an ABC model comparison approach. In this model comparison, an out-of-Africa replacement model was clearly favored over a multiregional evolution model. By analyzing the demographic history of both chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) in a combined ABC framework, Wegmann et al. (2010) gained valuable insights into the migration patterns between ancestral populations. The authors of this study also showed that central chimpanzees (*P. t. troglodytes*) experienced a stable demographic history and are therefore likely representing the ancient population within *Pan troglodytes*. In contrast, the other two subspecies (*P. t. schweinfurthii* and *P. t. verus*) showed signals of severe bottlenecks, likely associated with founder effects when these populations split off from the central population to colonize of new territories. Thalmann et al. (2011) used ABC methods on genetic data from a combination of historic and contemporary DNA samples of western gorillas (*Gorilla gorilla*). This approach allowed them not only to estimate the splitting time of the Cross River gorilla (*G. g. diehli*) population from the ancestral western gorilla population, but also allowed to determine that the interruption of gene flow between the two populations was recent (~420 years). Furthermore, the authors of this study found that the Cross River gorilla population went through a strong population decline, which they, given the recent dating of the start of the decline (~320 years), clearly attribute to anthropogenic pressure.

1.3.2. Genetic Marker Systems

Successful reconstruction of the demographic history of a species requires the use of multiple independent and selectively neutral genetic markers, because the coalescent process generating the observed genealogies is highly stochastic (Nichols 2001; Rosenberg & Nordborg 2002; Nielsen & Beaumont 2009). The genealogy of any given genetic marker will therefore contain information about true demographic processes as well as random noise, making it impossible to draw reliable conclusions about demographic processes based on a single genetic marker. Thus, multiple genetic markers must be used in demographic inferences. In addition, these genetic markers need to possess fully independent genealogical histories. This allows detecting true demographic signals by comparing multiple genealogies and identifying signals common to the genealogies of all employed genetic markers (Hare 2001; Rubinoff & Holland 2005). Neutral genetic markers have independent histories under a given demographic scenario if their recombination distance, that is the rate of recombination events between them, is approaching infinity (Rosenberg & Nordborg 2002). Thus, the genealogical histories of genetic markers located on the same chromosome are partly correlated, depending on their chromosomal distance, while markers on different chromosomes will always show completely independent genealogies within the constraints of

the common demographic history. It follows that all markers on the mitochondrial genome and the largest part of the Y chromosome (excluding the pseudo-autosomal region) have completely correlated histories, as recombination is absent or extremely rare (Tsaousis et al. 2005; Graves 2006). In the following paragraphs, I describe the main characteristics of the marker systems commonly used in phylogeographic studies and elaborate on their usefulness to reconstruct the population history of a species.

Genetic Marker Types

Three different types of genetic markers are mainly used in population genetic studies: sequence data, single-nucleotide polymorphisms (SNPs), and microsatellites (Sunnucks 2000). Sequence data is usually generated by amplifying and sequencing a specific stretch of DNA. All the sites within the same sequence are therefore highly correlated, but recombination events within the sequence stretch might occur, depending on the sequence length, the local recombination rate and the time to the most recent common ancestor of all sequences under consideration.

SNPs are variants of a single nucleotide, and such markers are usually widely distributed on the genome and therefore at least partially independent from each other. However, as a serious drawback, each single SNP contains only a very limited amount of information and reconstructing the genealogy of a single SNP with meaningful resolution is usually not possible, except in cases when SNPs can be grouped into sets of non-recombining haplotypes, as is the case for most Y-SNPs.

Microsatellites offer an interesting compromise between information content per locus and the possibility of using multiple independent markers with genome-wide distribution. Microsatellites are short sequence motives of 1-6 base pairs (bp) that are tandemly repeated (Tautz 1989). Microsatellites mutate by gaining or losing whole repeat units and usually have mutation rates up to five orders of magnitude higher than nucleotide substitutions (Schlötterer 2000), making them highly suitable for population studies (Sunnucks 2000). By amplifying a microsatellite locus by polymerase chain reaction (PCR) and measuring the length of the amplified fragment, the number of repeat units can be deduced. This allows analyzing microsatellite markers in heavily degraded DNA samples, as it is often necessary when working with non-invasively collected or ancient/historic DNA samples (Taberlet et al. 1996; Taberlet et al. 1999; Morin et al. 2001).

Sex-specific Marker Systems

Sex-specific markers systems are defined by their uniparental mode of inheritance, meaning that a genetic unit is only transmitted to the offspring by either the mother or the father. In most mammals, this applies to the mitochondrial genome with a strict maternal inheritance and the Y chromosome with a strict paternal inheritance. Since both the mitochondrial genome and the Y chromosome are largely free of recombination, such genetic marker systems allow reconstructing maternal and paternal lineages relatively easily. A further important characteristic of sex-specific marker systems is the difference in effective population size as compared to autosomal markers. Due to the haploidy and uniparental inheritance, both mitochondrial and Y-chromosomal markers have an effective population

size of one fourth of autosomal markers under the assumption of equal sex-ratios and equal variance in reproductive success of both sexes. In natural populations, the difference can be much higher, as many species show pronounced male reproductive skew (Trivers 1972).

1.3.3. DNA Sources

Obtaining DNA samples from evasive and endangered animals like orangutans is not a simple task. Due to legal restrictions and the need to avoid disturbance of animals, non-invasive sampling is often the only possible option to obtain DNA samples directly from wild populations (reviewed in Taberlet et al. 1999). Two sources of DNA are commonly used in such cases: fecal material and shed hair. In orangutans, collection of fecal material is only feasible if an individual has been followed until defecation, due to the difficulty of correctly identifying orangutan stool and the rapid degradation of host DNA in fecal material under tropical conditions. Thus, fecal samples are usually obtained during behavioral focal follows of orangutans within the range of long-term research sites. Furthermore, fecal samples require proper storage after collection, preferentially being frozen right after defecation to avoid excessive DNA degradation.

Even if storage conditions are optimal, only very small amounts of orangutan-specific DNA can be extracted from fecal samples, mixed with large amounts of bacterial and plant DNA (Taberlet et al. 1996; Morin et al. 2001). Moreover, the host DNA obtained from fecal material is usually highly degraded due to the bacterial activity in the stool material. This poses serious challenges and limitations for successful genetic analysis. First, polymerase chain reaction (PCR) primers are required to be highly specific to ensure amplification of solely the target DNA. Second, due to the small template amounts, PCR artifacts such as allelic drop-outs are common (Gerloff et al. 1995; Taberlet et al. 1996). Thus, in order to produce reliable genotypes from fecal material, multiple independent repetitions of each PCR are necessary (Taberlet et al. 1996; Morin et al. 2001). Third, the highly degraded DNA limits the size of the DNA stretch that can be amplified by PCR. Amplicon sizes of more than 500 bp can usually not be amplified consistently with template DNA from fecal material (Taberlet et al. 1999). This amplicon size limitation in combination with the common occurrence of allelic drop-outs would make it extremely expensive and time consuming to sequence large stretches of nuclear DNA. Rather, mitochondrial sequences and microsatellites are the preferred type of genetic markers when working with fecal material.

Shed hair is a very promising alternative DNA source in orangutans. Even though shed hair has analysis-wise similar limitations as fecal material, it can be obtained much easier from unhabituated animals outside established research sites, as it is often found in deserted night nests. Orangutans usually build a fresh nest every evening (van Schaik et al. 1995). Such nests are preserved in the forest for multiple months and can be easily identified. Acceptable DNA quality can be obtained from hair out of nests that are already a few days old, making it possible to obtain DNA samples in reasonable time even at locations that show extremely low population densities (<1 individual/km^2).

Given the limitations imposed by the low DNA quantity and quality of non-invasively collected samples, tissue or blood samples are the only way to acquire large amounts of nuclear sequence information. Such samples can only be obtained from either zoo individuals

or from rehabilitation centers, where confiscated or otherwise captured orangutans are prepared for their release into the wild. In both cases, blood samples are routinely taken during veterinary examination and thus are also available for DNA analysis, but the provenance of these individuals is often poorly recorded or unreliable. However, this limitation can be eased if a comprehensive genetic database of orangutan samples from the wild is available, which can then be used to assign captive individuals to their source population.

1.4. Aims and Organization of the Thesis

1.4.1. Motivation

This study aims to tackle the question of how different evolutionary processes have shaped the patterns of genetic variation in Sundaland, using orangutans as a study species. Of all great apes, the genetic make-up, population structure and demographic history of orangutans is clearly the least well understood, but potentially the most complex, giving the manifold geological processes at work during their evolutionary history. The lack of genetic studies dealing with these questions is remarkable, given the basal status of orangutans in the great ape lineage and the huge challenges that conservationists are facing with the genus. For several reasons, orangutans are very difficult to study. Due to their almost exclusively arboreal lifestyle, low population densities and the inaccessibility of their habitats, genetic sampling of a sufficiently large amount of individuals from a representative number of sites is extremely costly and time consuming. Furthermore, the legally protected status of orangutans makes it difficult to obtain sampling and export permits from local authorities and limits genetic sampling from wild population to non-invasive methods, which usually yield low quality and quantity DNA that are time consuming and costly to analyze (Morin et al. 2001).

The challenges associated with orangutan samples from wild populations have led researchers to extensively use samples from zoos and rehabilitation centers (Muir et al. 2000; Warren et al. 2001; Zhang et al. 2001; Kanthaswamy et al. 2006; Steiper 2006). The detailed provenance of these samples was often not known or unreliable, which made it impossible to discover small-scale geographic structuring of genetic diversity and lead to potentially wrong signals of gene flow. Furthermore, all these studies used only a small and geographically unrepresentative number of samples from Sumatran populations. Therefore, it was not possible to investigate the intriguing contrast of low census size and exceptionally high genetic diversity or the paraphyly of gene trees on Sumatra (Steiper 2006; Locke et al. 2011), which might shed light on the evolutionary history of the genus. Furthermore, our knowledge of the population structure and connectedness within the critically endangered Sumatran orangutan remains scarce and insufficient to inform conservation policies in any meaningful way.

Giving this critical lack of knowledge, this thesis aims to investigate the phylogeographic patterns and the population history of orangutans and identify relevant processes that led to population division and genetic differentiation. To achieve this, we pursued four main goals: First, we aimed to improve the coverage of genetic sampling from wild populations through an extensive collaborative effort with other field researchers and own genetic sampling.

Second, we investigated patterns of sex-specific dispersal on large spatial and temporal scales by employing both maternally and paternally inherited marker systems. Third, we inferred fine-scale population structure of both Bornean and Sumatran populations by employing highly variable autosomal microsatellite markers on a large number of samples per population, allowing for individual based approaches to questions of population structure and migration. Last, we aimed to apply novel modeling approaches to test specific scenarios of population history and obtain estimates for split times, migration rates and population size changes over multiple independent marker systems.

1.4.2. Outline

This study is structured into three data chapters, which I briefly outline in the following paragraphs. Each chapter represents a manuscript either already published in a peer-reviewed journal or to be submitted in the near future.

Chapter 2: Sex-biased Dispersal and Volcanic Activities Shaped Phylogeographic Patterns of Extant Orangutans (genus: Pongo)

In this study, published in Molecular Biology and Evolution (Nater et al. 2011), we investigated the geographic distribution and genealogical relationships of sex-specific genetic lineages in both orangutan species. This study improved our understanding of orangutan phylogeography relative to previous studies on two major points. First, due to an extensive collaborative effort with a large number of field researchers, we were able to analyze the most comprehensive set of wild samples with reliable provenance information to date. This allowed us to resolve fine-scale geographic structuring of genetic diversity throughout the current distribution of the genus. We applied up-to-date Bayesian methods to obtain estimates of divergence times of geographically defined lineages, which allowed us to infer potential processes that were responsible for major phylogenetic splits. Second, due to the use of both maternally inherited mitochondrial markers and paternally inherited Y-chromosomal markers, we could investigate sex-specific patterns of dispersal and detect to what extent geographical forces affect male and female orangutans differentially. We show that mitochondrial lineages on Sumatra are deeply geographically structured, with mtDNA lineages from a Sumatran site south of Lake Toba being more closely related to Bornean lineages. Mitochondrial lineages exhibited a coalescence time of 3.5 Ma, indicating not only the absence of female-mediated gene flow over the exposed Sunda shelf during Pleistocene glaciations, but also a strong separating effect of the Toba volcano. In contrast, the strong geographic clustering of mtDNA lineages was not present in the distribution of Y-chromosomal haplotypes, which also showed a much more recent coalescence (~170 kya), pointing at frequent long-distance dispersal of male orangutans.

Chapter 3: Marked Population Structure and Recent Migration in the Critically Endangered Sumatran Orangutan (Pongo abelii)

This chapter, published in the Journal of Heredity (Nater et al. 2013), investigates the population structure within Sumatran orangutans, an aspect that has seen very little efforts in the past, even though the species is critically endangered. We capitalized on a large sample set of wild orangutans from seven sampling regions on northern Sumatra, both sampled in the

wild and at rehabilitation centers. By analyzing mitochondrial HVRI sequences and autosomal microsatellite markers, we were able to reveal fine-scale population structure on Sumatra and evaluate the role of geographical features in shaping this structure. Furthermore, we investigated patterns of recent gene flow among genetic clusters to assess current habitat connectivity and detect important dispersal corridors that should be main targets of habitat protection efforts. Owing to the use of autosomal microsatellites, we were able to identify three distinct autosomal clusters within the limited range of Sumatran orangutans. These clusters were delimited by the caldera of the Toba volcano and the major Alas River. In contrast to the strong clustering of mitochondrial haplotypes, genetic separation was not complete on the autosomal level, as we identified signals of recent male-mediated migration between two genetic clusters north of Lake Toba. Furthermore, we were able to show that, despite heavy anthropogenic habitat degradation, most local subpopulations so far maintained relatively high levels of genetic diversity. Given this novel findings, this study is of high relevance for conservation efforts, which should aim to preserve the natural connectivity among local subpopulations in the critically endangered Sumatran orangutan.

Chapter 4: Reconstructing the Demographic History of Orangutans (Pongo spp.) using Approximate Bayesian Computation

In this study, we used a modeling approach based on Approximate Bayesian Computation (ABC) to test the fit of genetic data to specific models of orangutan demographic history. We aimed at enlarging our knowledge of the demographic history of orangutans by incorporating major improvements over previous studies dealing with this topic in four areas. First, we used an extensive set of genetic samples from wild orangutans with known provenance to adequately represent the genetic diversity present on both Borneo and Sumatra. Second, we included different genetic marker systems into a combined analysis to shed light on both female and male-specific histories. Third, we tested realistic demographic models based on the current knowledge of the geological history of Sundaland as well as the behavioral ecology and population genetics of orangutans. Last, in contrast to previous demographic modeling approaches, we also investigated complex demographic scenarios, including population structure, sudden population size changes and sex-biased gene flow patterns. By comparing demographic models with varying degrees of complexity, we demonstrate that a complex demographic model incorporating deep population structure and a recent decline on Sumatra as well as a bottleneck on Borneo during the last glacial period fits the currently observed genetic patterns in orangutans best. We also found evidence for regular and strongly male-biased migration between Borneo and Sumatra, which ceased at the beginning of the last glacial period. These results demonstrate that orangutans have experienced a complex demographic history, which was strongly influenced by the drastic environmental changes during the Pleistocene, and more recently by anthropogenic pressures.

Chapter 2:

Sex-biased Dispersal and Volcanic Activities Shaped Phylogeographic Patterns of Extant Orangutans (genus: *Pongo*)

Alexander Nater[1], Pirmin Nietlisbach[1,2], Natasha Arora[1], Carel P. van Schaik[1], Maria A. van Noordwijk[1], Erik P. Willems[1], Ian Singleton[3,4], Serge A. Wich[1,4], Benoit Goossens[5,6], Kristin S. Warren[7], Ernst J. Verschoor[8], Dyah Perwitasari-Farajallah[9,10], Joko Pamungkas[9,11], and Michael Krützen[1]

(1) Anthropological Institute & Museum, University of Zurich, Zurich, Switzerland
(2) Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Switzerland
(3) Foundation for a Sustainable Ecosystem (YEL), Medan, Indonesia
(4) PanEco, Foundation for Sustainable Development and Intercultural Exchange, Berg am Irchel, Switzerland
(5) Danau Girang Field Centre, c/o Sabah Wildlife Department, Kota Kinabalu, Malaysia
(6) Cardiff School of Biosciences, Cardiff University, Cardiff, United Kingdom
(7) School of Veterinary and Biomedical Sciences, Murdoch University, Perth, Australia
(8) Department of Virology, Biomedical Primate Research Centre, Rijswijk, The Netherlands
(9) Primate Research Center, Bogor Agricultural University, Bogor, Indonesia
(10) Department of Biology, Bogor Agricultural University, Bogor, Indonesia
(11) Department of Animal Infectious Diseases and Veterinary Public Health, Bogor Agricultural University, Bogor, Indonesia

Keywords: *Pongo* spp., Phylogeography, Sundaland, Dispersal, Glaciations, Toba

This chapter has been published in Molecular Biology and Evolution (2011) 28 (8): 2275-2288.

Author contributions:

Alexander Nater conceived the study, conducted genetic sampling on Sumatra, developed and performed laboratory procedures, conducted statistical analyses, and wrote the manuscript.

Carel P. van Schaik and Michael Krützen conceived the study and edited the manuscript.

Pirmin Nietlisbach performed laboratory procedures and reviewed the manuscript.

Natasha Arora performed laboratory procedures and reviewed the manuscript.

Erik P. Willems contributed map material.

Maria A. van Noordwijk, Ian Singleton, Serge A. Wich, Benoit Goossens, Kristin S. Warren, Ernst J. Verschoor, Dyah Perwitasari-Farajallah, and Joko Pamungkas provided genetic samples.

2.1. Abstract

The Southeast Asian Sunda archipelago harbors a rich biodiversity with a substantial proportion of endemic species. The evolutionary history of these species has been drastically influenced by environmental forces, such as fluctuating sea levels, climatic changes and severe volcanic activities. Orangutans (genus: *Pongo*), the only Asian great apes, are well suited to study the relative impact of these forces due to their well-documented behavioral ecology, strict habitat requirements and exceptionally slow life history. We investigated the phylogeographic patterns and evolutionary history of orangutans in the light of the complex geological and climatic history of the Sunda archipelago. Our study is based on the most extensive genetic sampling to date, covering the entire range of extant orangutan populations. Using data from three mitochondrial DNA (mtDNA) genes from 112 wild orangutans, we show that Sumatran orangutans, *Pongo abelii*, are paraphyletic with respect to Bornean orangutans (*P. pygmaeus*), the only other currently recognized species within this genus. The deepest split in the mtDNA phylogeny of orangutans occurs across the Toba caldera in northern Sumatra and, not as expected, between both islands. Until the recent past, the Toba region has experienced extensive volcanic activity, which has shaped the current phylogeographic patterns. Like their Bornean counterparts, Sumatran orangutans exhibit a strong, yet previously undocumented structuring into four geographical clusters. However, with 3.50 Ma, the Sumatran haplotypes have a much older coalescence than their Bornean counterparts (178 kya). In sharp contrast to the mtDNA data, 18 Y-chromosomal polymorphisms show a much more recent coalescence within Sumatra compared to Borneo. Moreover, the deep geographic structure evident in mtDNA is not reflected in the male population history, strongly suggesting male-biased dispersal. We conclude that volcanic activities have played an important role in the evolutionary history of orangutans and potentially of many other forest-dwelling Sundaland species. Furthermore, we demonstrate that a strong sex bias in dispersal can lead to conflicting patterns in uniparentally inherited markers even at a genus-wide scale, highlighting the need for a combined usage of maternally and paternally inherited marker systems in phylogenetic studies.

2.2. Introduction

The Southeast Asian Sunda archipelago harbors a rich biodiversity with a substantial proportion of endemic species, including 21 nonhuman primates (Harrison et al. 2006). This exceptional diversity has its roots in the special geological history of the Sundaland region. The phylogenetic patterns that we observe today within and among terrestrial species living on Sundaland were formed by four main forces:

First, tectonic plate movements opened up and destroyed land bridges between islands. This allowed a multitude of plant and animal species to colonize the archipelago from the Southeast Asian mainland (Meijaard 2004), but subsequently separated the island populations from the source populations. In the case of Sundaland, tectonic plate movements might have been responsible for phylogenetic splits older than 2.5 Ma, but ceased to play a role when the Sunda islands reached their present shape in the Early Pleistocene (Meijaard 2004).

The second major force is recurring glacial periods, which influenced the Sundaland species in that falling sea levels led to the temporary exposure of the Sunda shelf (Lisiecki & Raymo 2005), repeatedly uniting the Sundaland islands into a single landmass. The underlying climate changes affected the extent and type of vegetation on the exposed landmass in Sundaland (Morley 2000; Cannon et al. 2009). During glacial periods, the climate was generally cooler, drier and more seasonal. Under such conditions, evergreen rainforest would have been restricted to mountain slopes and areas in proximity to the coast, whereas the central plains on the emerged shelf may have been dominated by savannah and grassland (Gathorne-Hardy et al. 2002; Bird et al. 2005). The glacial periods were therefore either an isolating or a connecting force, depending on the species concerned and its habitat requirements. Irrespective of the specific effect, glaciations were a recurrent force that repeatedly influenced the evolutionary histories of species since the Late Pliocene (Harrison et al. 2006).

The third major force shaping phylogenetic patterns were rivers. Large rivers pose insuperable barriers for many terrestrial species (e.g. Ayres & Clutton-Brock 1992). Yet, the precise characteristics that determine the strength of the barrier, that is, course, width, depth and flow rate, are subject to dynamic changes, both seasonal and long-term. Additionally, while rivers present barriers to gene flow across them, the accompanying forest galleries on either side might have also acted as important dispersal corridors for forest-dwelling species during dry periods (Gorog et al. 2004).

Last, volcanic eruptions, the fourth force, may have played a substantial role in the evolutionary history of species in Sundaland. The Southeast Asian region has an eventful history of volcanic eruptions (Hall 1996), which may have led to local extinctions and subsequent recolonizations of areas devastated by volcanic activities. In addition, the recent Toba supereruption, which occurred around 73 kya on northern Sumatra and is considered to be one of the most powerful volcanic eruptions in geological history (Rose & Chesner 1987; Chesner et al. 1991; Williams et al. 2009), is thought to have had significant consequences for the flora and fauna on Sundaland and potentially worldwide.

Among the extant endemic Sundaland species, orangutans are ideally suited to study the relative impact of these four forces thanks to their well-documented behavioral ecology (e.g.

Delgado & Van Schaik 2000; Wich et al. 2009b). Their habitat is restricted to evergreen rainforest due to a pronounced arboreality and primarily frugivorous diet (Delgado & Van Schaik 2000). Behavioral studies suggest that orangutans show female philopatry and male dispersal (Galdikas 1995; Singleton & van Schaik 2002; van Noordwijk & van Schaik 2005; Morrogh-Bernard et al. 2011; van Noordwijk et al. 2012). Both sexes of orangutans generally do not cross rivers when opposite banks are not connected via the canopy (Rijksen & Meijaard 1999). The exceptionally slow life history of orangutans (Wich et al. 2009a) is also useful in detecting genetic signals of old demographic events. Furthermore, due to low densities of orangutans (Wich et al. 2004), effective population sizes are expected to be small, which in turn will lead to fast lineage sorting.

Orangutans had an eventful evolutionary history, as shown by the drastic changes in distribution over the last few million years (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). A rich subfossil record bears testimony to the enormous area that ancestral populations once inhabited, spanning from southern China to northeast India, mainland Southeast Asia and most of Sundaland (Delgado & Van Schaik 2000). The current range, however, is highly restricted, with extant populations only found in forest patches on northern Sumatra (*Pongo abelii*) and on Borneo (*P. pygmaeus*) (Wich et al. 2008). Climatic changes during the Pleistocene as well as anthropogenic factors, such as hunting by prehistoric hunter-gatherer societies, have been suggested as causes for this drastic collapse in the ancient orangutan distribution range (Jablonski 1998; Delgado & Van Schaik 2000). More recently, extensive habitat destruction has been stated as the main cause for the massive population decline that occurred during the last two centuries (Goossens et al. 2006a).

Sumatran and Bornean orangutans are regarded as separate species, even though they produce fertile hybrid offspring in captivity (Muir et al. 1998). The cause of this incomplete reproductive isolation might be the slow divergence due to the exceptionally long generation time of orangutans compared with other mammals (Wich et al. 2009a) or, alternatively, recent gene flow and in consequence a mixing of gene pools between the islands. Indeed, two studies (Muir et al. 2000; Kanthaswamy et al. 2006) identified mitochondrial haplotypes in Sumatran orangutans, clustering with haplotypes commonly found on Borneo. Furthermore, a study describing simian foamy viruses (SFV) found certain Sumatran virus lineages to be more closely related to those found in Bornean orangutans rather than to Sumatran ones (Verschoor et al. 2004).

Another interesting characteristic relating to orangutans as compared with other Sundaland primates is the exceptionally high level of genetic diversity of the Sumatran species (Zhi et al. 1996; Muir et al. 2000; Steiper 2006; Thinh et al. 2010). This has been explained through the recolonization of the island from the mainland, Borneo and Java following the Toba supereruption (Muir et al. 2000). However, because many of the samples used in previous studies (Muir et al. 2000; Zhang et al. 2001; Kanthaswamy et al. 2006; Steiper 2006) came from captive individuals with unknown provenance, support for such a complex scenario requires much more detailed information about the population structure within Sumatra.

This study uses an unprecedented dense sampling regime of wild orangutans from both Sumatra and Borneo with sufficiently detailed provenance to investigate the phylogeographic history of this genus. We noninvasively collected samples from ten Bornean and six Sumatran

sites, covering most of the current range of both Bornean and Sumatran orangutans (Figure 2.1). We complemented our dataset with samples from rehabilitation centers, provided location of capture from the wild was known. This improved population coverage, especially for regions where collection of samples posed major logistic and bureaucratic challenges. Additionally, we used two differentially transmitted marker systems (mitochondrial DNA [mtDNA] and Y-chromosomal), which enabled us to assess the effects of sex-biased long-range dispersal in orangutans.

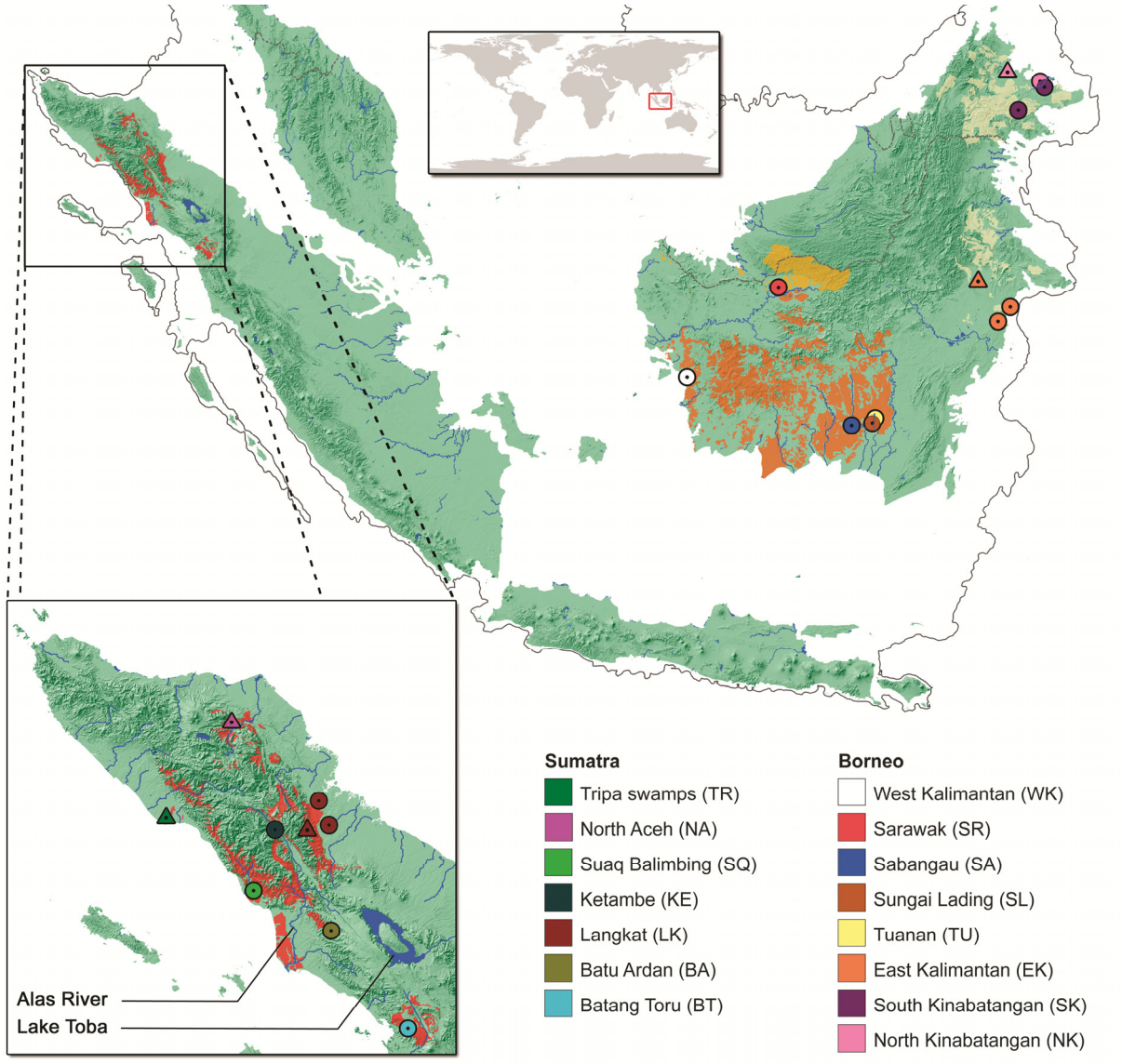


Figure 2.1: Sampling locations in Sumatra and Borneo. The colored areas represent the current distribution of the Sumatran orangutan and the three Bornean subspecies. Circles denote sampling from wild populations, whereas triangles represent the origin of rehabilitant orangutans. The contour of the Sunda shelf during the last glacial maximum (120 m below current sea level) is indicated by a thin black line.

2.3. Materials and Methods

2.3.1. Sample Collection

The orangutan samples used in this study were obtained from three different sources. We collected fecal samples noninvasively at various field sites from wild orangutan populations across the entire range of both species. On Sumatra, this included the areas of Suaq Balimbing, Ketambe, Sikundur, Sampan Getek, Batu Ardan and Batang Toru. In Borneo, we obtained samples from Gunung Palung, Sabangau, Sungai Lading, Tuanan, Danum Valley, and north and south sides of the Lower Kinabatangan (Figure 2.1). We also obtained blood and hair samples from confiscated wild-born orangutans at rehabilitation centers, including the Sumatran Orangutan Conservation Program in Medan, Wanariset in East Kalimantan, Semongok in Sarawak and Sepilok in Sabah. Last, we included hair samples that had been collected by KW (Warren et al. 2001) from wild orangutans, either from nests or plucked during translocation, in East Kalimantan (Kutai and Sangatta) and Northwest Kalimantan (Danau Sentarum) (Figure 2.1). Supporting Table S2.1 lists all the samples used for this study.

Fecal samples were either stored in RNAlater Storage Solution (Applied Biosystems) or 90% EtOH, or collected in 90% EtOH and later dried with silica gel (Nsubuga et al. 2004). Hair samples were placed in paper envelopes and dried with silica gel. Blood samples were taken during routine veterinary examination and collected in standard EDTA blood collection tubes. All samples were stored at -20°C as soon as possible.

The collection and transport of samples were conducted in strict accordance with Indonesian, Malaysian and regulations of the international community. Samples were transferred to Zurich under the Convention on International Trade in Endangered Species (CITES) from Indonesia (permits 09717/IV/SATS-LN/2010, 07279/IV/SATS-LN/2009, 00961/IV/SATS-LN/2007, 06968/IV/SATS-LN/2005), Sabah, Malaysia (permit 4645) and the United Kingdom (reexport permit 290569/01).

2.3.2. DNA Extraction

Fecal samples were extracted using a QIAamp DNA Stool Mini Kit on a QIAcube robotic workstation (both Qiagen) following the standard extraction protocol for human DNA extraction from stool samples with elution in 100 µl AE buffer (Qiagen). Blood samples were processed with a QIAamp DNA Blood Mini Kit (Qiagen) according to manufacturer's instructions and eluted in 100 µl AE buffer. Hair samples have been extracted using an EZ1 DNA Investigator Kit on a BioRobot EZ1 Workstation (both Qiagen), applying the pretreatment for DNA extraction from hair samples as described in the Investigator Kit manual with elution in 100 µl Tris-EDTA buffer.

2.3.3. PCR Amplification, Sequencing and Genotyping

Identity Analysis

We confirmed the unique identity of samples collected in the wild by genotyping all extracts with six highly polymorphic microsatellite markers (D2S141, D5S1505, D6S501, D13S321,

D13S765, and D16S420) following the procedure as described in Arora et al. (2010). When matching genotypes were found, only one sample was included in the following analyses.

Mitochondrial DNA

For this study, we analyzed the three mtDNA loci 16S ribosomal DNA (16S), Cytochrome b (CYTB) and NADH-ubiquinone oxidoreductase chain 3 (ND3). For the sequencing of the 16S locus we used the primers from Zhi et al. (1996), and for CYTB and ND3 primers from Muir et al. (2000). All polymerase chain reactions (PCRs) were performed in a 10 µl volume containing 1 µl genomic DNA, 0.2 U HotStarTaq DNA Polymerase, 1x PCR buffer (both Qiagen) containing 1.5 mM final concentration $MgCl_2$, 0.1 mM dNTPs and 0.1 µM each of forward and reverse primer. PCR amplifications were performed in a Veriti Thermal Cycler (Applied Biosystems) with the following parameters: initial denaturation at 95°C for 15 min, 35 cycles (blood extracts) or 45 cycles (fecal and hair extracts) of 94°C for 30 s, 61°C for 40 s and 72°C for 40 s, followed by a final extension step at 72°C for 10 min. PCR products were examined on a 1.5% agarose gel and the product yield was estimated by comparing with a reference band with known concentration in a 100 bp DNA ladder (New England BioLabs). Cycle sequencing was performed in a 10 µl reaction containing 1–3 ng of unpurified PCR product, 1x sequencing buffer (80 mM Tris, 2 mM $MgCl_2$, pH 9.0), 0.4 µM forward primer and 0.3 µl BigDye Terminator v3.1 on a 3730 DNA Analyzer (both Applied Biosystems). We sequenced the reverse strand only in cases where the quality or length of the forward strand sequence was unsatisfactory for reliable base calling. The SEQMAN program of the LASERGENE 8 software package (DNASTAR) was used to trim and align the sequences. None of our sequences contained an insertion or deletion. The sequences are deposited on GenBank under the accession numbers HQ912716–HQ912752.

Y Polymorphisms

We genotyped Y-linked microsatellites and single nucleotide polymorphisms (SNPs) by using fluorescently labeled forward primers. For the microsatellites, we sequenced multiple reference alleles for each marker to infer the number of microsatellite repeats in relation to the fragment length. For SNP typing, we used the labeled forward primer in combination with two interrogating reverse primers, differing in their last 3'-base and primer length. This yielded PCR products differing by a few bases depending on SNP state. We combined these two methods to type SNPs in microsatellite flanking regions together with the repeat length of the microsatellite. This allowed us to genotype 6 Y-linked SNPs, one insertion-deletion polymorphism and 11 microsatellites in two multiplex PCR reactions as described elsewhere (Nietlisbach et al. 2010).

The PCR products were diluted 20–85 times with ddH_2O and 1 µl was mixed with 9.95 µl HiDi formamide and 0.07 µl GeneScan 500 LIZ size standard (both Applied Biosystems). After 3 min of denaturation at 95 °C the PCR products were run on a 3730 DNA analyzer and analyzed with GENEMAPPER v4.0 (both Applied Biosystems).

2.3.4. Phylogenetic Analyses

We calculated summary statistics and genetic differentiation measures for mtDNA sequence data and Y-chromosomal polymorphisms with ARLEQUIN v3.5.1.2 (Excoffier & Lischer

2010). The same software was used for the analysis of molecular variance (AMOVA). For the mtDNA data set, we concatenated the three loci to a single sequence of 1,355 bp and applied a Tamura and Nei distance correction (Tamura & Nei 1993) with a gamma value of 0.281, as determined by jMODELTEST v0.1.1 (Posada 2008) to be the best-fitting model supported by ARLEQUIN. For the Y-chromosomal data set, we treated the single nucleotide and insertion/deletion polymorphisms like biallelic microsatellites and used a sum of squared differences based distance measurement (R_{ST}-like) for all Y-chromosomal polymorphisms (Slatkin 1995).

To keep the number of comparisons small and increase the sample size, we pooled the following sampling sites and origins of rehabilitant orangutans into broader sampling regions: Sikundur, Sampan Getek and rehabilitant orangutans from the Langkat region into Langkat (LK), Gunung Palung and rehabilitant animals from West Kalimantan into West Kalimantan (WK), Danau Sentarum and rehabilitant orangutans from Semongok into Sarawak (SR), Kutai, Sangatta and rehabilitant animals from East Kalimantan into East Kalimantan (EK), Danum Valley and south side of the Lower Kinabatangan River into South Kinabatangan (SK), and north side of Kinabatangan together with the rehabilitant orangutans from Sepilok into North Kinabatangan (NK). This grouping is justified by mtDNA haplotype sharing within these sampling regions and non-significant differentiation of Y haplotypes (data not shown).

We constructed a phylogenetic tree for the three mtDNA loci using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST v1.6.1 (Drummond & Rambaut 2007). The phylogenetic tree was rooted with a human and a central chimpanzee sequence from GenBank (accession nos. GQ983109.1 and HM068590.1, respectively). The BEAST software was also used to estimate divergence dates under a relaxed molecular clock model with uncorrelated lognormal distributed branch substitution rates (Drummond et al. 2006). We chose a birth-death speciation process to generate the prior distribution of node ages (Yang & Rannala 2006). Two nodes were used to calibrate the molecular clock by defining a normal prior distribution for the node age: 1) the *Pan*/*Homo* divergence with a mean age of 6.5 Ma and a standard deviation of 0.3 (Brunet et al. 2002; Vignaud et al. 2002); 2) the *Pan-Homo/Pongo* divergence with a mean age of 18.3 Ma and a standard deviation of 3.0 (Steiper & Young 2006). We defined a very broad prior distribution for the second calibration point to take into account the differing divergence dates found in other studies (see Raaum et al. 2005 for an overview).

Because our dataset incorporated an rRNA sequence, protein-coding regions and flanking transfer RNA sequences, we tested different data partitioning schemes, from one single partition up to eight partitions (rRNA, tRNA, and 1st, 2nd and 3rd codon position of CYTB and ND3, respectively). We used jMODELTEST to determine the nucleotide substitution model of each partition (Supporting Table S2.2). We performed BEAST test runs for each partitioning scheme over 20,000,000 generations starting from a random tree and sampling every 1,000 generations (Supporting Table S2.3). For each test run, we used TRACER v1.5 (Rambaut & Drummond 2007) to obtain the marginal likelihood, that is, the probability of the data given the tested model. TRACER uses a harmonic mean estimator on the MCMC likelihood trace together with bootstrapping to calculate the marginal likelihood and its

standard error (Newton et al. 1994; Redelings & Suchard 2005). We preferred a more complex model over a simpler model if the ratio of their corresponding marginal likelihoods, that is, the Bayes Factor, was larger than 100 (Supporting Table S2.4, Jeffreys 1961; Kass & Raftery 1995).

Given the marginal likelihoods of every partitioning scheme, we decided to partition our data into four partitions (all RNA sequences and all coding sequences, coding sequences subdivided into 1st, 2nd and 3rd codon position). A TN93+I (Tamura & Nei 1993) nucleotide substitution model was selected by jMODELTEST as the minimal adequate models under the Bayesian information criterion (Schwarz 1978) for all partitions. We unlinked nucleotide frequencies, substitution rates, and proportion of invariable sites for each partition, but combined all partitions into a single tree topology. The MCMC analysis was run four times independently for 20,000,000 generations each, starting from a random tree and sampling every 1,000 generations. After completing all runs, we used TRACER to examine run convergence. We aimed for an effective sample size of at least 200 for all parameters, which was the case after discarding the first 20% of samples as burn-in and then combining the 16,000 samples of each run with LOGCOMBINER v1.6.1 (part of the BEAST software package). TREEANNOTATOR v1.6.1 (part of the BEAST software package) was used to draw a maximum clade credibility tree. Tree visualization was done in FIGTREE v1.3.1. We performed an additional BEAST run with an empty alignment but otherwise identical settings in order to get a sample from the prior distributions of all parameters and node ages.

Due to the generally accepted problems of using microsatellite markers for phylogenetic inference (Wilson & Balding 1998; Richard & Thorpe 2001), we used the BATWING software (Wilson et al. 2003) for Y-linked markers, as it employs a coalescent-based approach within a Bayesian framework to infer a gene tree of all Y-chromosomal haplotypes together with parameter estimates for coalescence times and effective population size. We used a constant size population model with all Bornean and Sumatran orangutans combined into two single populations, respectively. This is justified because of the Y-haplotype sharing we found among study sites within each island. We coded four SNPs as unique event polymorphisms (inftype=0) because they were diagnostic for a clear distinction between both species and appeared to be monomorphic within. The other SNPs showed a more complex mutational history and were excluded from the analysis. We allowed each microsatellite locus to have its own mutation rate for which we defined a gamma prior [3.75, 2,500] based on human studies as reviewed in Macpherson et al. (2004). The starting tree was obtained by a parsimony heuristic search with a low random component (badness=0.01). We performed four independent MCMC runs taking 10,000 samples each, with 1,000 parameter changes and 20,000 tree swaps between each sampling occasion. After removing 20% of the samples from each run as burn-in, we combined the parameter outputs of all four runs in TRACER. We used the program PRIOR (part of the BATWING package) with the same settings as in the previous runs to sample the prior distributions of all parameters. To obtain the estimates for the coalescence times, we multiplied the corresponding coalescent units with the estimated effective population size for each sampled state. We assumed a generation time of 25 years for orangutans (Wich et al. 2009a) in order to convert the number of generations into years.

We used R version 2.12.1 (R Development Core Team 2010) to visualize prior and posterior distributions of both BEAST and BATWING runs and calculate summary statistics.

2.3.5. Median-joining Networks

Median-joining networks (Bandelt et al. 1999) using either the whole Y-chromosomal dataset (SNPs, insertion-deletion polymorphism, and microsatellites) or all 144 polymorphic positions in the three mtDNA loci were constructed using NETWORK v4.5.1.6 and NETWORK PUBLISHER v1.2.0.0 (http://www.fluxus-engineering.com). All Y-chromosomal loci were weighted by 100 times their gene identity (Nei 1987): $J=\Sigma x_i^2$ with x_i being the frequency of the i^{th} allele in the whole set of samples. This way, loci with lower mutation rates are weighted higher, which reduces the generation of noise due to homoplasy of faster evolving loci. We compared networks with high and low epsilon values, but as high epsilon values and thus more reticulations did not show obvious character conflicts, an epsilon value of zero was used for the networks presented.

2.4. Results

2.4.1. Mitochondrial and Y-chromosomal Haplotype Networks

Both mtDNA and Y-chromosomal haplotypes networks show a clear separation of Bornean and Sumatran haplotypes (Figure 2.2). Yet, both networks reveal pronounced marker-specific differences in the distribution of haplotypes within each island. The mtDNA network exhibits a remarkably strong geographic structure, especially within Sumatra. Surprisingly, however, the southernmost extant Sumatran population of Batang Toru connects to the Bornean orangutans rather than to the other Sumatran sites. In contrast to this, the only clear geographic clustering in the Y-haplotype network is between Borneo and Sumatra. Within each island, Y haplotypes are shared among sites and no geographic structure is visible by eye. The haplotypes found in Batang Toru cluster with the other Sumatran Y haplotypes and do not connect to the Bornean haplotypes as seen in the mtDNA network. In summary, the structural differences found in the haplotype networks of mtDNA and Y chromosomes point toward a strongly male-biased dispersal in orangutans, especially with respect to long-distance movements.

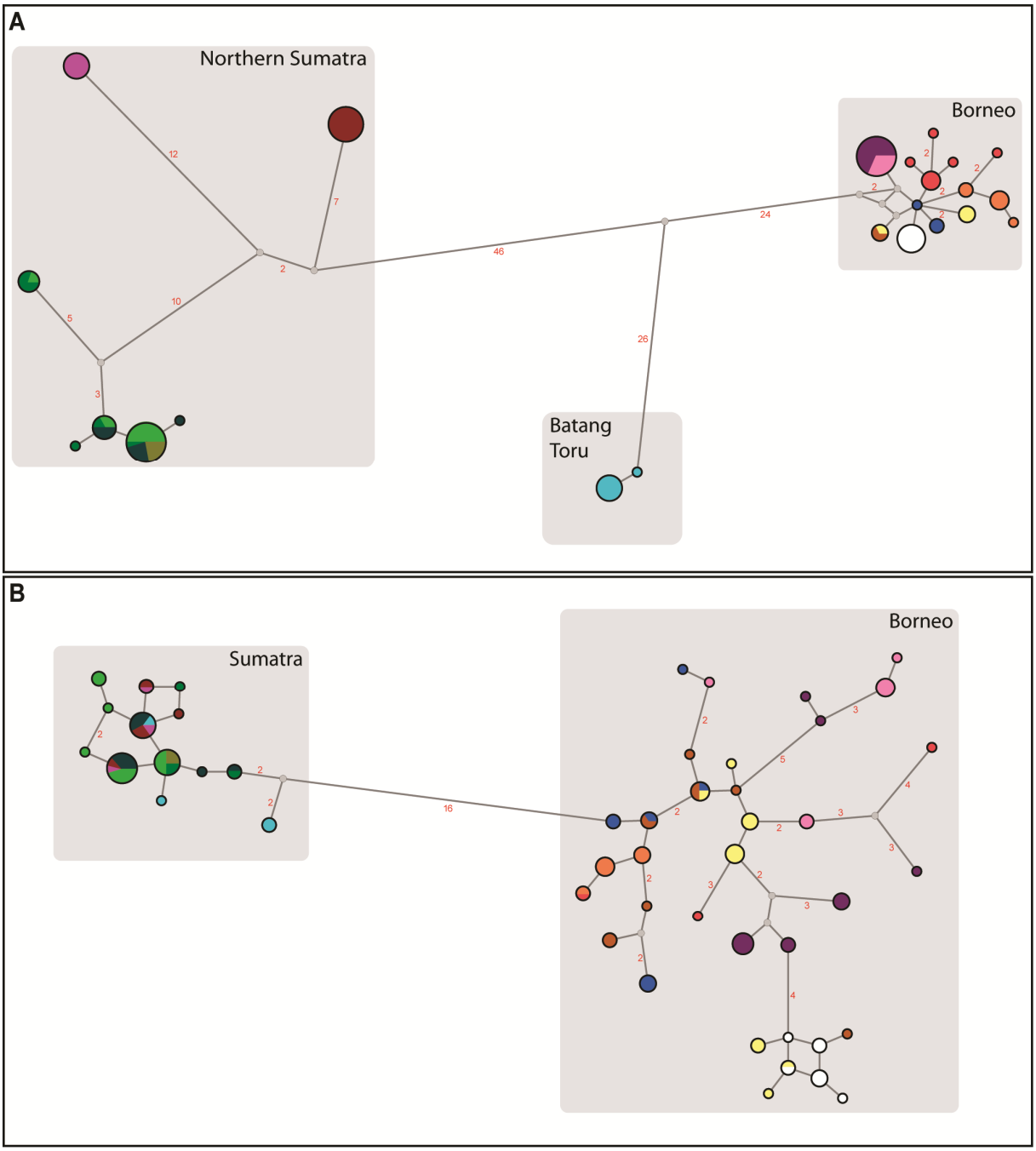


Figure 2.2: Median-joining networks of (A) mtDNA and (B) Y-chromosomal haplotypes. The red numbers in between the nodes indicate connections with more than one mutational step. The node colors represent the different sites, as indicated in Figure 2.1. The size of each node is proportional to the number of individuals with identical haplotypes.

2.4.2. Mitochondrial and Y-chromosomal Diversity and Differentiation

As evident from the haplotype networks, the two marker systems show opposite patterns of genetic diversity between Borneo and Sumatra (Table 2.1). For mtDNA data, Sumatran orangutans have a ten times higher mean pairwise difference among haplotypes (π) compared with Borneans. For the Y-chromosomal haplotypes, the mean sum of squared differences in Borneo is nearly five times the size of the Sumatran species.

Table 2.1: Summary statistics for all examined orangutan sampling regions

	mtDNA			Y-polymorphisms		
Sample set[a]	$N_{Samples}$	MPD[b]	HD[c]	$N_{Samples}$	SSD[d]	HD[c]
All *Pongo*	112	47.338	0.913	115	36.192	0.973
Sumatra	60	30.553	0.824	40	3.299	0.862
Tripa (TR)	7	5.143	0.715	4	3.500	0.833
North Aceh (NA)	7	0.000	0.000	3	4.667	1.000
Suaq (SQ)	12	1.742	0.439	13	2.590	0.781
Ketambe (KE)	8	0.786	0.679	9	2.333	0.751
Langkat (LK)	14	0.000	0.000	5	2.800	0.900
Batu Ardan (BA)	4	0.000	0.000	2	0.000	0.000
Batang Toru (BT)	8	0.250	0.250	4	5.667	0.833
Borneo	52	2.971	0.829	75	15.515	0.976
West Kalimantan (WK)	9	0.000	0.000	11	1.691	0.909
Sarawak (SR)	8	2.250	0.786	3	26.000	1.000
Sabangau (SA)	3	0.667	0.666	8	8.500	0.857
Sungai Lading (SL)	2	0.000	0.000	10	9.289	0.933
Tuanan (TU)	4	2.000	0.500	13	4.282	0.871
East Kalimantan (EK)	7	0.762	0.667	8	0.786	0.679
S. Kinabatangan (SK)	13	0.000	0.000	14	13.571	0.791
N. Kinabatangan (NK)	6	0.000	0.000	8	16.179	0.750

Sumatran and Bornean sites are shaded in light and dark grey, respectively. [a], The sampling regions of LK, EK, and SK contain multiple sampling sites (see section 2.3.4. for details); [b], Mean pairwise nucleotide difference; [c], Haplotype diversity (Nei 1987); [d], Sum of squared allele size differences.

Genetic differentiation analyses using Φ_{ST} (mtDNA) and R_{ST} (Y polymorphisms) support a strong and significant separation of Bornean and Sumatran orangutans, but the two marker systems exhibit different levels of differentiation among sites for each species (Table 2.2). In Borneo, both marker systems show generally high levels of differentiation among sites. Interestingly, the populations across the Kinabatangan River are significantly differentiated for the Y haplotypes, but show only a single mtDNA haplotype. The Sumatran sites are much more differentiated at the mtDNA than the Y-chromosomal level, as all pairs of sites, with the exception of Suaq Balimbing, Ketambe and Batu Ardan, are significantly differentiated for mtDNA. In sharp contrast to the strong mtDNA structure, the Y-chromosomal markers indicate a panmictic population in Sumatra.

Table 2.2: Pairwise population differentiation values for mtDNA (Φ_{ST}, above diagonal) and Y-chromosomal (R_{ST}-like, below diagonal) markers

Φ_{ST}/R_{ST}	TR	NA	SQ	KE	LK	BA	BT	WK	SR	SA	SL	TU	EK	SK	NK
TR	-	0.90*	0.40*	0.47*	0.93*	0.44*	0.98*	0.98*	0.97*	0.96*	0.96*	0.96*	0.97*	0.98*	0.97*
NA	-0.31	-	0.96*	0.99*	1.00*	1.00*	1.00*	1.00*	0.99*	1.00*	1.00*	0.99*	1.00*	1.00*	1.00*
SQ	0.20	0.13	-	-0.04	0.97*	-0.09	0.99*	0.99*	0.98*	0.99*	0.99*	0.98*	0.99*	0.99*	0.99*
KE	-0.01	0.02	0.10	-	0.99*	0.09	1.00*	1.00*	0.99*	0.99*	0.99*	0.99*	0.99*	1.00*	1.00*
LK	-0.18	-0.33	0.26	0.18	-	1.00*	1.00*	1.00*	0.99*	1.00*	1.00*	1.00*	1.00*	1.00*	1.00*
BA	-0.26	-0.38	-0.22	-0.37	-0.05	-	1.00*	1.00*	0.99*	1.00*	1.00	0.99*	1.00*	1.00*	1.00*
BT	0.25	0.20	0.19	0.15	0.40	-0.05	-	1.00*	0.98*	1.00*	1.00*	0.99*	0.99*	1.00*	1.00*
WK	0.97*	0.97*	0.97*	0.97*	0.97*	0.98*	0.96*	-	0.63*	0.91*	1.00*	0.80*	0.92*	1.00*	1.00*
SR	0.77*	0.75	0.89*	0.86*	0.82*	0.69	0.74*	0.67*	-	0.32*	0.54*	0.44*	0.64*	0.77*	0.67*
SA	0.88*	0.88*	0.92*	0.90*	0.90*	0.87*	0.85*	0.75*	-0.05	-	0.84	0.46	0.80*	0.96*	0.93*
SL	0.88*	0.88*	0.91*	0.90*	0.89*	0.86*	0.85*	0.63*	0.03	0.02	-	0.53	0.87*	1.00*	1.00*
TU	0.93*	0.93*	0.94*	0.94*	0.94*	0.93*	0.92*	0.56*	0.24	0.44*	0.23*	-	0.76*	0.89*	0.81*
EK	0.97*	0.97*	0.96*	0.97*	0.97*	0.99*	0.95*	0.94*	0.20	0.04	0.28*	0.73*	-	0.95*	0.92*
SK	0.79*	0.79*	0.86*	0.83*	0.81*	0.78*	0.80*	0.53*	0.10	0.41*	0.44*	0.35*	0.56*	-	0.00
NK	0.88*	0.88*	0.92*	0.91*	0.90*	0.86*	0.87*	0.46*	0.44*	0.56*	0.44*	0.48*	0.71*	0.59*	-

Comparisons within northern Sumatra and Borneo are shaded in light and dark gray, respectively. Comparisons between the islands are in medium gray, and comparisons involving Batang Toru are in white.

* $p<0.05$.

2.4.3. Analysis of Molecular Variance

Both marker systems show a distinct differentiation between the two species/islands (Table 2.3). For Y-chromosomal loci, 82% of the total molecular variance is partitioned between Bornean and Sumatran orangutans. For mtDNA, species differences explain 75% of the total variance. However, if we define Batang Toru (the only extant Sumatran population south of Toba) as a third group, among-group variance grows to over 89%. At the species level, the AMOVA confirmed our findings regarding different levels of population differentiation between mtDNA and Y-chromosomal markers. On Borneo, 50% of the Y-chromosomal variation is partitioned among sites versus 81% for mtDNA. This difference is even more pronounced for Sumatran orangutans with 12% versus 98% for Y-chromosomal and mtDNA loci, respectively.

Table 2.3: Results of AMOVA for both marker systems and four geographical partitions

	mtDNA (TN93+Γ[a])		Y chromosome (SSD[b])	
Geographical partitions	Variance	% Variance	Variance	% Variance
Sumatra and Borneo (N_{Groups}=2)				
Among groups	37.68*	74.54	26.77*	81.56
Among populations, within groups	12.43*	24.58	2.77*	8.45
Within populations	0.45*	0.89	3.28*	10.00
Sumatra, Batang Toru, and Borneo (N_{Groups}=3)				
Among groups	45.47*	89.42	25.34*	80.44
Among populations within groups	4.93*	9.70	2.88*	9.14
Within populations	0.45*	0.88	3.28*	10.42
Within Sumatra (N_{Groups}=1)				
Among populations	22.03*	97.54	0.21	12.14
Within populations	0.56*	2.46	1.48	87.86
Within Borneo (N_{Groups}=1)				
Among populations	1.39*	81.36	4.12*	49.69
Within populations	0.32*	18.64	4.17*	50.31

[a], Tamura and Nei distance method with gamma correction; [b], sum of squared size differences; *, $p<0.01$

2.4.4. Phylogeny and Molecular Dating

Using Bayesian phylogenetic methods, we found very strong support for a strict separation between Bornean and Sumatran orangutan populations for both mtDNA and Y-chromosomal loci (posterior probability 1.00), with no haplotype sharing occurring between the two islands. However, for mtDNA, the population of Batang Toru is more closely related to the Bornean orangutans than the other Sumatran sites (posterior probability 1.00). The maximum clade credibility trees for the mtDNA and Y-chromosomal data sets are shown in Figure 2.3.

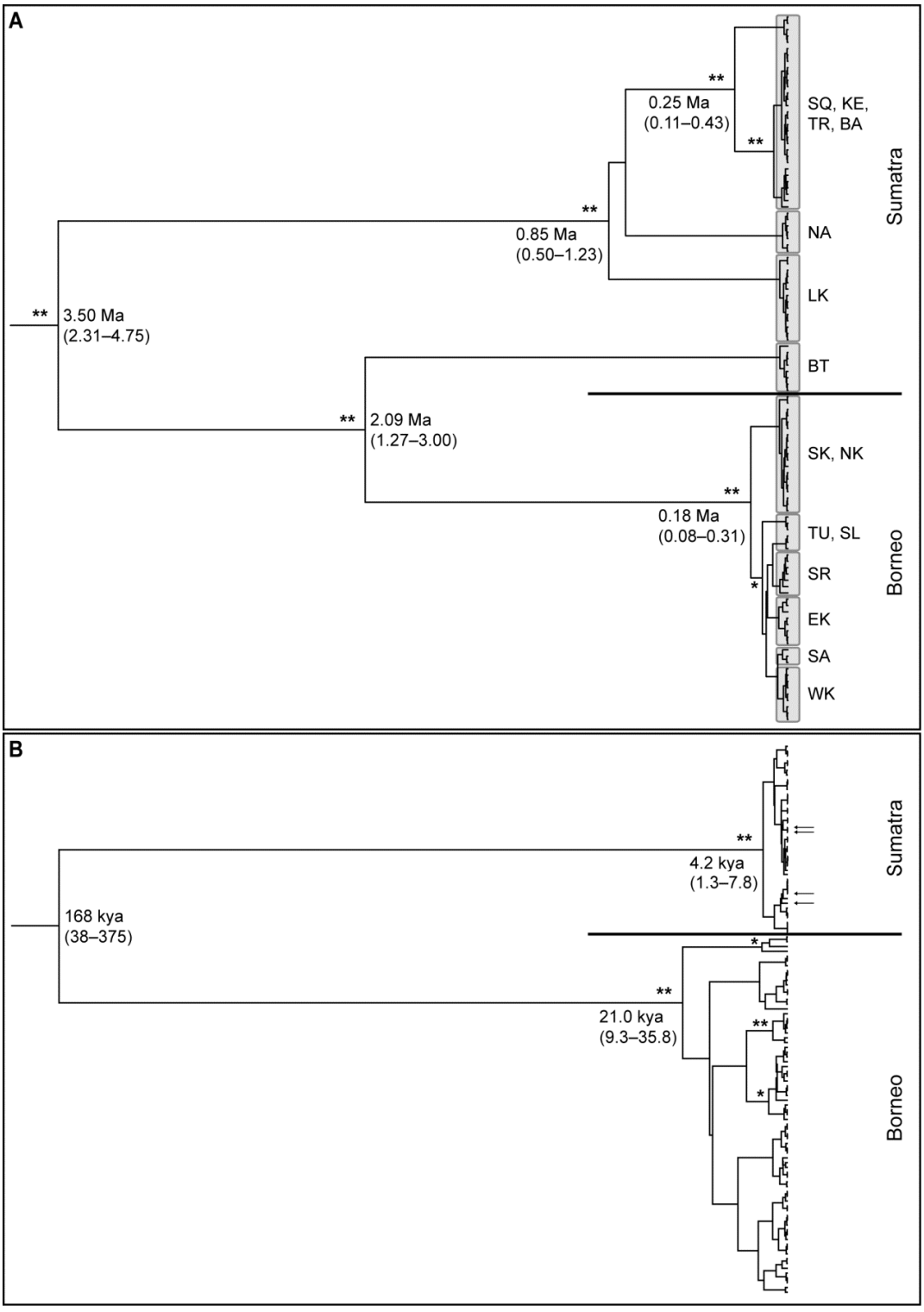


Figure 2.3: (A) Consensus tree of concatenated mtDNA loci from the posterior sample of the BEAST analysis. The gray shaded boxes indicate the different sampling regions. The tree is rooted with a human and a chimpanzee sequence as outgroup (not shown). (B) Consensus tree of Y-chromosomal loci from the posterior sample of the BATWING analysis. The arrows within the Sumatran cluster mark the samples from Batang Toru. The posterior probability of each clade is indicated with stars above the basal node: **, >95%; *, 50–80%. Nodes within the shaded areas are not annotated due to space constrains. The node ages are mean values of the posterior probability distribution and are given together with the 95% highest probability density interval.

Due to the position of Batang Toru in the rooted tree, Sumatran orangutans form a paraphyletic group. They exhibit a strong clustering into four geographically distinct groups, all supported by a posterior probability of 1.00. The groups correspond to the areas of Batang Toru, Langkat, a cluster made up of only rehabilitant individuals that likely originated from North Aceh, and what seems to be a large population spanning from Batu Ardan in the southeast over Suaq Balimbing to Tripa in the northwest and Ketambe in the east. No mtDNA haplotype sharing occurs among the different clusters.

Bornean orangutans on the other hand form a monophyletic group with a very recent divergence. The sites in Sabah (Danum Valley, south and north side of Lower Kinabatangan) form an outgroup to all other Bornean regions (posterior probability 0.61).

In contrast to previous studies (Muir et al. 2000; Kanthaswamy et al. 2006), we found no evidence for recent, that is, during the last glacial maximum, migration events between Borneo and Sumatra. According to our molecular dating approach, the haplotype lineages of Batang Toru on one side and all Bornean populations on the other side were separated 2.09 Ma (95% highest posterior density interval: 1.27–3.00 Ma). The radiation within Borneo occurred around 178 kya (75–305 kya). The two lineages giving rise to all Bornean orangutans together with the population in Batang Toru on one side and all the other Sumatran orangutans on the other side separated 3.50 Ma (2.31–4.75 Ma). Within Sumatra, the next oldest divergence is 0.85 Ma (0.50–1.23 Ma), which is still significantly older than what we could observe within Borneo. The posterior distributions of the divergence time estimates are shown in Figure 2.4A together with the corresponding prior distributions.

The molecular dating with BATWING on our Y-chromosomal dataset resulted in coalescence dates that were much more recent than the values obtained from mtDNA. We estimated the coalescence date for the Bornean orangutans to be 21.0 kya (95% highest posterior density interval: 9.3–35.8 kya) with an effective population size of 1,083 (603–1,623). For the Sumatran orangutans we obtained 4.2 kya (1.3–7.8 kya) and an effective population size of only 180 (62–327) Y chromosomes. We estimated the time to the most recent common ancestor (TMRCA) of all orangutan Y chromosomes at 168 kya (38–375 kya), 20 times more recent than the TMRCA based on mtDNA. This value was based on an estimated effective population size of 1213 (661–1,815 years) male orangutans and is robust in terms of the predefined population structure and population growth models. Only the prior distribution for the microsatellite mutation rates had a large impact on the coalescence date estimates. The prior and posterior distributions of the BATWING estimates are shown in Figure 2.4B.

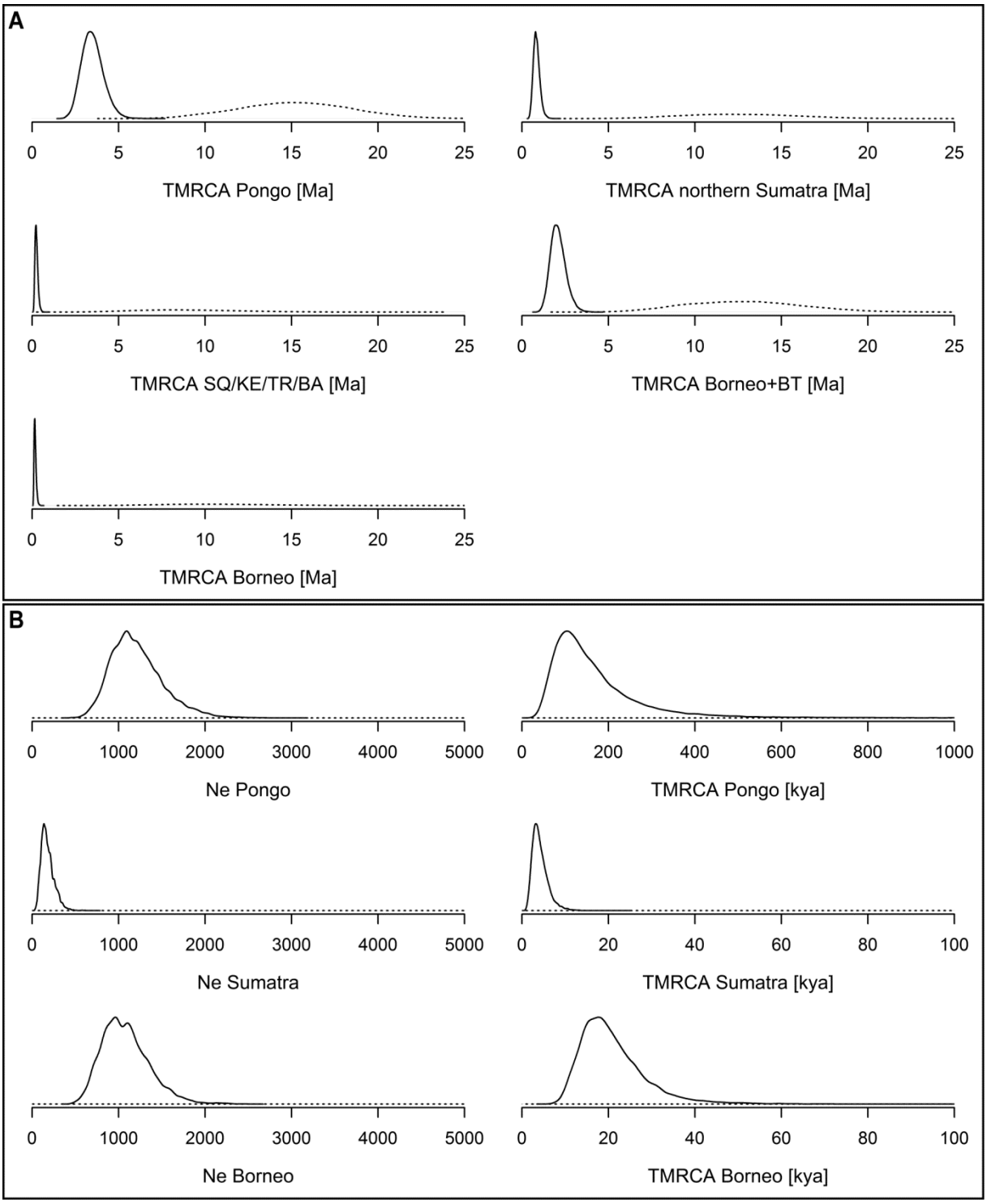


Figure 2.4: Posterior (solid lines) and prior (dotted lines) distributions for the relevant parameter estimates in (A) BEAST (mtDNA) and (B) BATWING (Y-chromosomal loci).

2.5. Discussion

Owing to the use of male- and female-mediated marker systems, we were able to draw the most complete picture of the evolutionary history of the only Asian great ape to date. The most unexpected finding is that at the mtDNA level the southernmost Sumatran population of Batang Toru forms a different clade from the rest of Sumatra with much higher affinity to the

Bornean populations. The existence of this clade is surprising given that the geographic distance from Batang Toru to the nearest Sumatran site in our sample is a mere 140 km. The mtDNA lineage found in Batang Toru diverged around 3.5 Ma from the lineage leading to all other Sumatran mtDNA haplotypes and separation appears to have been maintained since then. The separation line corresponds with the position of Lake Toba, a gigantic caldera complex that originated from a sequence of at least four major and countless smaller volcanic eruptions within the last 1.2 Ma (Chesner et al. 1991). The Batang Toru population seems to be the remnant of a large South and West Sumatran gene pool. This gene pool also gave rise to the lineage leading to all extant Bornean orangutans around 2.1 Ma, as indicated by the basal position of Batang Toru to the Bornean cluster.

The area of Lake Toba is known to represent a significant zoogeographic boundary for many bird species (Whitten et al. 2000), but also primates. Both the Thomas' leaf monkey (*Presbytis thomasi*) (Aimi & Bakar 1996) and white-handed gibbon (*Hylobates lar*) (Whittaker et al. 2007; Thinh et al. 2010) are in Sumatra restricted to areas north of Lake Toba, whereas the mountain agile gibbon (*Hylobates agilis*) (Whittaker et al. 2007; Thinh et al. 2010) occurs in Sumatra only south of Lake Toba. It is conceivable that pyroclastic flows and lava streams completely destroyed the rainforest between the caldera and both the Malacca strait and the Indian Ocean (Ambrose 2003). This would have repeatedly sealed off northern Sumatra from the rest of the island for all forest-dependent species. Under such circumstances, the surroundings of Toba are expected to form a strong dispersal barrier for rainforest species like orangutans and gibbons. Once forest contact was restored, the populations would meet again, but not merge, due to the female orangutans' philopatric tendencies (Galdikas 1995; Singleton & van Schaik 2002; van Noordwijk & van Schaik 2005; Morrogh-Bernard et al. 2011).

In contrast to our findings from mtDNA, the Y-chromosomal network shows that the strong matrilineal separation across Lake Toba is not paralleled by the male orangutan history. Male-driven gene flow appears to have prevented the divergence of the populations north and south of the Toba barrier into distinct taxonomic units, despite the presence of a strong separation line in the female population history. Male orangutans, in contrast to females, have been observed migrating through forest types and at altitudes unlikely to provide enough food for long-term survival (Rijksen & Meijaard 1999). Such wanderers are presumed to be males that have not established a stable home range. These orangutans seem to be able to temporarily tolerate less productive habitats in their search for a more suitable forest patch not currently occupied by a dominant male. Long-distance movements of these males in combination with successful reproduction outside their natal area will have facilitated the exchange of Y-chromosomal haplotypes between Batang Toru and northern Sumatra.

Contrary to previous studies (Muir et al. 2000; Verschoor et al. 2004; Steiper 2006), there is no evidence for recent gene flow between Borneo and Sumatra. Both mtDNA and Y-chromosomal data confirm a long-lasting separation between the two islands. We attribute such signals in other studies to human translocation events or unreliable sample provenance, a possibility also acknowledged by the authors of these studies themselves.

The absence of recent gene flow between both islands might at first seem surprising given the cyclical exposure of the Sunda shelf during the Pleistocene associated with land bridges between Borneo and Sumatra as recent as 10,000 years ago (Voris 2000; Inger & Voris 2001).

However, during low sea levels, major river systems dissected the exposed shelf (Rijksen & Meijaard 1999; Harrison et al. 2006). Moreover, several studies suggest that glaciation periods were characterized by a drier and more seasonal climate (Morley 2000; Bird et al. 2005). Under such circumstances, the forests on the land bridges were most likely separated by broad savannah corridors (Bird et al. 2005; Harrison et al. 2006). This is supported by similar patterns of deep divergence of island specific mtDNA lineages in other forest-dwelling species in Sundaland, for example, murine rodents (Gorog et al. 2004), Sunda pig-tailed macaques (Ziegler et al. 2007), and gibbons (Thinh et al. 2010).

Our results are not consistent with any of the previously reported scenarios of orangutan population history. The deep divergence of mtDNA lineages and high genetic diversity within Sumatra had already been noted in previous studies (Muir et al. 2000; Kanthaswamy et al. 2006; Steiper 2006). Given the approximately ten times smaller current census size of Sumatran versus Bornean orangutans (Wich et al. 2008), it was interpreted as a strong signal for a complex demographic history. Steiper (2006) suggested that Sumatran orangutans either represent the remains of a large contiguous population spanning from the Malay Peninsula over Sumatra to Java or that they are made up of individuals originating from multiple differentiated ancient populations that migrated to northern Sumatra in the past. Muir et al. (2000) hypothesized that such a population admixture was linked to the Toba supereruption around 73 kya, which might have eradicated all populations on Sumatra and promoted a recolonization of the island from mainland Southeast Asia, Borneo and Java.

The first scenario of a large panmictic population is, however, inconsistent with our finding of a deep geographic structure in the distribution of mtDNA haplotypes, with nearly 98% of the mtDNA variation in Sumatra partitioned among the sampling locations. The same geographic structuring also makes the second scenario of a Sumatran population composed of multiple immigrant lineages unlikely. Isolation mechanisms must have been in place to prevent the admixture of differentiated mtDNA lineages following immigration to northern Sumatra. However, given such separating forces, the deep mtDNA divergence within Sumatra is most parsimoniously explained by a stable population history with deep substructuring due to persistent migration barriers and highly restricted dispersal of female orangutans (Galdikas 1995; Singleton & van Schaik 2002; van Noordwijk & van Schaik 2005).

Previous studies have shown that rivers are strong migration barriers for Bornean orangutans (Goossens et al. 2005; Jalil et al. 2008; Arora et al. 2010). Interestingly, however, in our study we did not find a significant population differentiation for either mtDNA or Y-chromosomal markers between the sites of Batu Ardan and Ketambe or Suaq Balimbing, which are separated by the major Alas River. This suggests that the Alas is only a weak migration barrier in the downstream areas, probably because it forms clear meanders in its lowland stretch, which are occasionally cut off (Gascon et al. 2000).

In contrast to mtDNA, the Y haplotypes in Sumatra show little geographic structure and an extremely recent coalescence. We explain this discrepancy by pronounced sex differences in dispersal behavior, and a small male effective population size in Sumatran orangutans. Contrary to females, unflanged male orangutans seem to leave the area of birth and may often need to cover large distances in order to establish a new home range (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000; Singleton & van Schaik 2001). Hence, the Y-

chromosomal network of orangutans reflects little geographic structure. A similar but opposite pattern of highly different levels of geographic structuring of mtDNA and Y-chromosomal haplotypes has been found in chimpanzees (Langergraber et al. 2007) and bonobos (Eriksson et al. 2006). These results match the expectations from behavioral studies that point toward male philopatry and female dispersal in these species (e.g. Goodall 1983; Kano 1992), further supporting the interpretation of our results. The coalescence estimate for the Sumatran Y chromosomes of only slightly over 4,000 years and the corresponding effective population size of less than 200 Y chromosomes seems unexpectedly low. However, such a small male effective population size, and therefore recent Y-chromosomal coalescence, makes sense in the light of the drastic population decline that started in the Holocene and continues until today (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000; Goossens et al. 2006a). Additionally, the pronounced male dominance hierarchy in Sumatran orangutans, and consequent reproductive skew, might drastically reduce the number of reproducing males (Setia & van Schaik 2007; Utami Atmoko et al. 2009b).

In summary, glaciations and associated sea level changes alone cannot explain the phylogenetic patterns observed in orangutans. The deepest split in the mtDNA phylogeny occurs within the island of Sumatra and, not as expected, between Sumatra and Borneo. This lineage division does not correspond to the course of any major river. Tectonic plate movements could also not have played a role in the separation of mtDNA lineages on Sumatra, as the island acquired its current shape during the Early Pleistocene (Meijaard 2004), which would have left sufficient time to homogenize the gene pools on Sumatra. We therefore propose that volcanic activities explain the permanent separation of mtDNA lineages between the populations north and south of Lake Toba best. The further subdivision into three geographically distinct mtDNA clusters north of Lake Toba can be explained by a combination of river effects and socio-behavioral dispersal barriers.

Apart from the separation by the Toba barrier, Sumatran orangutan populations were not severely affected by the Toba eruptions. The paraphyly of Sumatran mtDNA lineages, along with the deep geographically anchored splits in the populations north of Toba, indicates a remarkably strong temporal stability of local orangutan populations, and thus the forests they inhabit, despite the seemingly devastating impact of the Toba explosions. This surprising stability of orangutan populations close to the caldera throws some doubt on reconstructions suggesting wholesale habitat destruction over large regions following the supereruption (Rampino & Ambrose 2000; Williams et al. 2009). The Sumatran pattern stands in stark contrast to Borneo with its comparatively extremely recent divergence of mtDNA lineages (~180 kya), which points toward a strong refugium on the island (Arora et al. 2010).

Our results for orangutans, as well as previous work on other Sundaland species (e.g. Gorog et al. 2004; Meijaard & Groves 2004; Thinh et al. 2010), clearly demonstrate that Pleistocene sea level changes alone are not sufficient to explain the evolutionary history of forest-dwelling species on the Sunda archipelago. The strong impact of volcanic activities as evidenced in the phylogeographic patterns of orangutans most certainly also played a significant role in the evolutionary history of many other Sundaland species. Finally, we show that sex differences in dispersal not only influence the genetic makeup of local populations, but can shape genetic relationships on a species or even genus-wide scale. Our results

highlight the need for male and female-inherited markers systems to obtain powerful and more complete insights into the evolutionary history of a species.

2.6. Acknowledgements

We thank Ellen Meulman, Andrea Permana, Helga Peters, Gregoire Bertagnolio, Gail Campbell-Smith, Yenny Saraswati, Rachmad Wahyudi, Meredith Bastian, Cheryl Knott, Helen Morrogh-Bernard, Noko Kuze and Tomoko Kanamori for collecting samples in the field or in rehabilitation centers. Nadja Morf extracted some of the samples used in this study. Maja Greminger, Michael E. Steiper, and three anonymous reviewers gave valuable comments on the manuscript. This study was financially supported by the Swiss National Science Foundation (grant no. 3100A-116848 to MK and CvS), Messerli Foundation, A.H.-Schultz Foundation and Claraz Schenkung. Furthermore, we thank the following institutions for supporting our research: Indonesian State Ministry of Research and Technology (RISTEK), Indonesian Institute of Sciences (LIPI), Sabah Wildlife Department, Taman National Gunung Leuser (TNGL), Borneo Orangutan Survival Foundation (BOSF), Leuser International Foundation (LIF), and Badan Pengelola Kawasan Ekosistem Leuser (BPKEL).

Chapter 3:

Marked Population Structure and Recent Migration in the Critically Endangered Sumatran Orangutan (*Pongo abelii*)

Alexander Nater[1], Natasha Arora[1], Maja P. Greminger[1], Carel P. van Schaik[1], Ian Singleton[2,3], Serge A. Wich[1,3], Gabriella Fredriksson[2,3], Dyah Perwitasari-Farajallah[4,5], Joko Pamungkas[4,6], and Michael Krützen[1]

(1) Anthropological Institute & Museum, University of Zurich, Zurich, Switzerland

(2) Foundation for a Sustainable Ecosystem (YEL), Medan, Indonesia

(3) Sumatran Orangutan Conservation Programme (PanEco-YEL), Medan, Indonesia

(4) Primate Research Center, Bogor Agricultural University, Bogor, Indonesia

(5) Department of Biology, Bogor Agricultural University, Bogor, Indonesia

(6) Department of Animal Infectious Diseases and Veterinary Public Health, Bogor Agricultural University, Bogor, Indonesia

Keywords: Great apes, gene flow, microsatellites, conservation, Sundaland

This chapter has been published in the Journal of Heredity (2013) 104 (1): 2-13.

Author contributions:

Alexander Nater conceived the study, conducted genetic sampling on Sumatra, developed and performed laboratory procedures, conducted statistical analyses, and wrote the manuscript.

Carel P. van Schaik and Michael Krützen conceived the study and edited the manuscript.

Natasha Arora and Maja P. Greminger performed laboratory procedures and reviewed the manuscript.

Ian Singleton, Serge A. Wich, Gabriella Fredriksson, Dyah Perwitasari-Farajallah, and Joko Pamungkas provided genetic samples.

3.1. Abstract

A multitude of factors influence how natural populations are genetically structured, including dispersal barriers, inhomogeneous habitats, and social organization. Such population subdivision is of special concern in endangered species, as it may lead to reduced adaptive potential and inbreeding in local subpopulations, thus increasing the risk of future extinctions. With only 6,600 animals left in the wild, Sumatran orangutans (*Pongo abelii*) are among the most endangered, but also most enigmatic, great ape species. In order to infer the fine-scale population structure and connectivity of Sumatran orangutans, we analyzed the most comprehensive set of samples to date, including mitochondrial hypervariable region I haplotypes for 123 individuals and genotypes of 27 autosomal microsatellite markers for 109 individuals. For both mitochondrial and autosomal markers, we found a pronounced population structure, caused by major rivers, mountain ridges, and the Toba caldera. We found that genetic diversity and corresponding long-term effective population size estimates vary strongly among sampling regions for mitochondrial DNA, but show remarkable similarity for autosomal markers, hinting at male-driven long-distance gene flow. In support of this, we identified several individuals that were most likely sired by males originating from other genetic clusters. Our results highlight the effect of natural barriers in shaping the genetic structure of great ape populations, but also point toward important dispersal corridors on northern Sumatra that allow for genetic exchange.

3.2. Introduction

Most natural populations do not behave like single units, in which random mating occurs over the entire distribution (Kimura & Weiss 1964). Rather, most populations are genetically structured, the extent of which is determined by several factors. Geographical factors include both isolation by distance (Wright 1943) and physical barriers impeding gene flow across them, such as mountain ridges, rivers, and deserts. Ecological factors concern the distribution of resources and predators, which may lead to an aggregation of individuals within high-quality habitat patches (Slatkin 1987). A third category includes social, mating and dispersal behaviors. Gregarious species, where individuals live in social groups, often show a marked population structure even in the complete absence of obvious geographical or ecological factors (Storz 1999; Ross 2001). Yet, strong genetic structuring imposed by limited dispersal has also been found in non-gregarious species. This is because in both gregarious and non-gregarious species it is potentially advantageous for individuals to show some degree of philopatry, as in the natal area food resources are familiar and kin is available for social support (Johnson & Gaines 1990; Handley & Perrin 2007). Moreover, dispersal is usually heavily biased towards one sex, because one major benefit of dispersal is the avoidance of inbreeding (Bengtsson 1978; Pusey 1987). As a consequence, the extent of observed genetic structure may vary greatly depending on the inheritance mode of the genetic marker system used to investigate such patterns.

The underlying genetic structure of populations is especially important from a conservation perspective. Genetic structure may lead to local isolation of gene pools, resulting in effective subpopulation sizes that are only a fraction of the effective population size in a population without substructure (Charlesworth 2009). This has three important evolutionary consequences. First, lower effective sizes of subpopulations lead to stronger genetic drift effects and a reduced number of mutation events in each subpopulation. As a consequence, genetic diversity within each subpopulation will be lower compared to that of an unstructured population. Moreover, deleterious mutations that would be eliminated by background selection in unstructured populations might become fixed in small subpopulations, thus reducing the average population fitness (Hedrick & Kalinowski 2000; Reed & Frankham 2003). Second, population structuring increases the chance of mating among relatives, therefore causing potential loss of fitness due to inbreeding depression (Hedrick & Kalinowski 2000). Third, local separation of genetic variants will allow different selection pressures to act on specific subpopulations, thus allowing for adaptations to specific local environmental conditions (Williams 1966; Kawecki & Ebert 2004). While local adaptations raise the average fitness of subpopulations in a constant environment, the loss of genetic diversity reduces the potential of the subpopulations to adapt to changing environmental conditions and therefore carries greater risks of future extinctions (Reed & Frankham 2003). All these negative effects, however, can be counterbalanced by gene flow among subpopulations (Slatkin 1987). Therefore, knowledge about the extent to which genetic diversity is structured and exchanged across the range of a species is crucial to predict the long-term survival of populations and to implement effective conservation measures.

Population subdivision is a major concern in large-bodied animals with small population sizes, slow life histories and low rates of reproduction, as such taxa are especially vulnerable to the aforementioned negative effects of population fragmentation (Hedrick & Kalinowski 2000). Great apes are of special interest in investigating the causes and consequences of population subdivision, not only because studying their population histories can reveal valuable insights into the evolution of modern humans, but also because all extant species are listed as endangered or even critically endangered (IUCN 2012). Furthermore, great apes show variation in dispersal patterns, which affects the genetic structuring of populations. For instance, chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) show female-biased dispersal (Tautz et al. 1999; Mitani et al. 2002), whereas males are the dispersing sex in orangutans (*Pongo* spp.) (Singleton & van Schaik 2002; Morrogh-Bernard et al. 2011; Arora et al. 2012; van Noordwijk et al. 2012), as is the case in most mammals (Dobson 1982). In contrast, in gorillas (*Gorilla* spp.), both sexes disperse, even though mean dispersal distance is different between males and females (Douadi et al. 2007).

In the past, a substantial body of work has investigated population structure in great apes, such as in chimpanzees (Becquet et al. 2007; Gonder et al. 2011), bonobos (Eriksson et al. 2004; Eriksson et al. 2006), gorillas (Bergl & Vigilant 2007; Guschanski et al. 2008), and Bornean orangutans (*Pongo pygmaeus*) (Warren et al. 2001; Goossens et al. 2005; Jalil et al. 2008; Arora et al. 2010). Yet, a detailed population genetic analysis of Sumatran orangutans (*Pongo abelii*) is still lacking, even though Sumatran orangutans are critically endangered (IUCN 2012). As of today, only an estimated 6,600 individuals remain in the wild, as compared to about 54,000 Bornean orangutans (Wich et al. 2008). In contrast to the Bornean species, where three subspecies have been defined based on morphological characters (Groves 2001), no subspecies have been proposed for Sumatran orangutans.

Historically, Sumatran orangutans populated most of the Indonesian island of Sumatra, as evidenced by fossil finds and historical records (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). The current distribution is, however, restricted to small forest patches on the northern tip of Sumatra (Wich et al. 2008). Ecological and anthropogenic factors, such as prehistoric hunting and recent deforestation, have been suggested as explanations for the drastic range collapse of orangutans (Delgado & Van Schaik 2000). The comparatively limited range of Sumatran orangutans that remains today is subdivided by major rivers and mountain ridges. Moreover, the massive forest exploitation that started in the last century (Rijksen & Meijaard 1999) has caused severe habitat fragmentation, leaving habitat blocks of continuous forest that often harbor only a few hundred individuals (Wich et al. 2008). This habitat fragmentation in combination with the potentially very strong reproductive skew in Sumatran orangutan males (Setia & van Schaik 2007; Utami Atmoko et al. 2009b) might have drastically reduced the effective sizes of local subpopulations, thus minimizing genetic diversity and posing a severe threat of future extinctions.

Sumatran orangutans show the strictest arboreality among all great apes (Delgado & Van Schaik 2000) and occur in two different rain-forest habitat types. Low-altitude peat-swamp forests offer high and constant food supplies and support the highest population densities (Husson et al. 2009). At lower densities, permanent populations of Sumatran orangutans can be found in dry-land forests up to an altitude of 1500 meters above sea level or more (Wich et

al. 2004; Husson et al. 2009). However, in non-riverine dry-land forests, the mast fruiting phenomenon causes extreme temporal fluctuations in food availability (Knott 1998; Husson et al. 2009), which may act as a strong selective pressure for adaptive traits related to prolonged food scarcity. Unfortunately, due to the absence of long-term field studies covering the entire extant range of Sumatran orangutans, little is known about variation in behavior, physiology and morphology within this species that could hint at the presence of habitat specific adaptations.

The current lack of knowledge about the genetic structure of Sumatran orangutans is mainly caused by difficulties in obtaining samples with reliable provenance throughout the entire species' range. This factor prevented most previous genetic studies from interpreting the extraordinary high diversity on the mitochondrial DNA (mtDNA) level they found in Sumatran orangutans as compared to their Bornean sister species (Muir et al. 2000; Kanthaswamy et al. 2006; Steiper 2006). However, using samples with a well-defined geographic origin, Nater et al. (2011) showed that mitochondrial variation is strongly geographically structured on Sumatra. This study identified four distinct mitochondrial clusters in Sumatran orangutans, with divergence times of up to 3.5 million years. Similar, albeit less pronounced patterns of geographical structuring of mtDNA has been found in Bornean orangutans (Warren et al. 2001; Arora et al. 2010). However, mtDNA is not a good indicator of population structure and gene flow in species that show a strong male-bias in dispersal, like orangutans (Galdikas 1995; Singleton & van Schaik 2002; Morrogh-Bernard et al. 2011; Arora et al. 2012; Nietlisbach et al. 2012; van Noordwijk et al. 2012). In fact, using Y-chromosomal markers, Nater et al. (2011) showed that the deep divergence and strong geographic clustering observed with mtDNA is not present in the male population history, indicating long-distance migration by males across Sumatra. The amount of gene flow and the resulting extent of homogenization of autosomal gene pools among local subpopulations is, however, impossible to measure using only sex-linked marker systems.

In this study, we aimed to unravel patterns of genetic diversity and differentiation in Sumatran orangutans, using a combination of mitochondrial and autosomal genetic markers. We investigated the role of geographical, ecological and behavioral factors underlying the fine-scale population structure and tested for connectivity among subpopulations. To achieve this, we analyzed the most comprehensive and largest set of orangutan samples from Sumatra to date, using samples from wild individuals originating from the entire species' range.

3.3. Materials & Methods

3.3.1. Sample Collection

Three different kinds of orangutan samples were analyzed for this study: First, fecal samples were collected non-invasively at long-term study sites. Second, in areas where animals were not habituated, we collected hair samples from deserted nests. Third, we obtained blood and hair samples of confiscated wild-born orangutans from the quarantine station of the Sumatran Orangutan Conservation Program (SOCP) in Medan, North Sumatra.

We obtained orangutan samples from seven different sampling regions (Figure 3.1A): Tripa (TR), North Aceh (NA, north of Tamiang River), West Leuser (WL), Central Leuser (CL,

west side of Alas River), Langkat (LK, east of Alas River, south of Tamiang River), Batu Ardan (BA, east of Alas River, west of Lake Toba), and Batang Toru (BT, south of Lake Toba) (Supporting Table S3.1). Fecal and hair samples were collected and stored following the genetic sampling protocol of the orangutan network (http://www.aim.uzh.ch/orangutannetwork). All blood samples were taken during routine veterinary examination in the SOCP quarantine station. Blood samples were collected in standard EDTA blood collection tubes and stored at -20°C.

The amount and reliability of information about the wild origin of rehabilitant orangutans varied considerably. We classified the provenance of these individuals as reliable if the location of confiscation was known in detail and if this location was near an extant wild orangutan population. The samples from rehabilitant orangutans that did not meet these criteria were classified as having unknown provenance and excluded from certain analyses (see below).

The collection and transport of samples was carried out in compliance with Indonesian and international regulations. Samples were exported from Indonesia to Zurich under the Convention on International Trade in Endangered Species (CITES) (permits 09717/IV/SATS-LN/2010, 07279/IV/SATS-LN/2009, 00961/IV/SATS-LN/2007, 06968/IV/SATS-LN/2005).

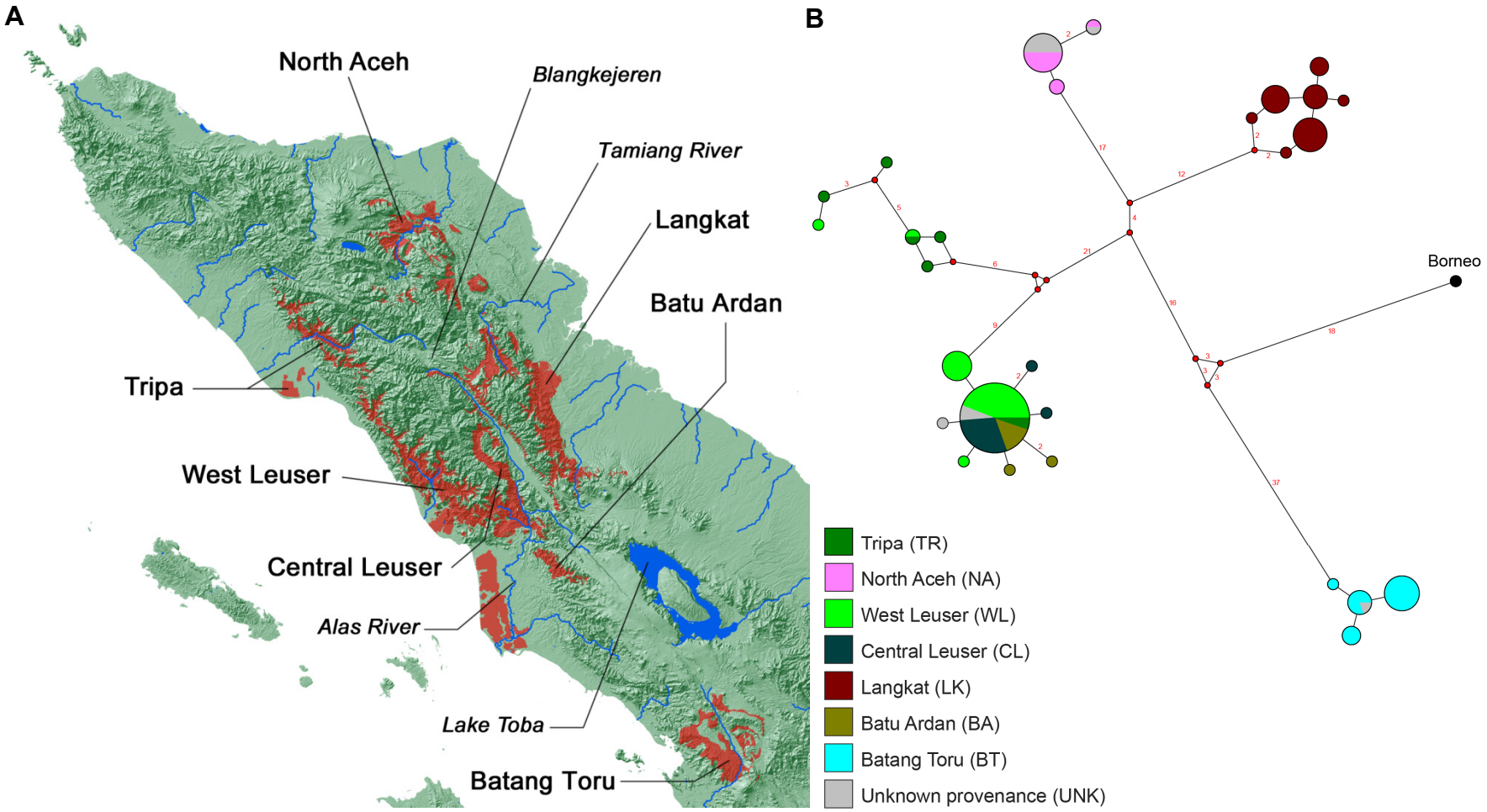


Figure 3.1: (A) Map of sampling regions in northern Sumatra. Labels in italics denote important geographic features. The red shading represents the current distribution of Sumatran orangutans. (B) Median-joining network of mitochondrial HVRI haplotypes. The red numbers in between the nodes indicate the number of mutational steps in between haplotypes (one step if not indicated otherwise). The size of each node is proportional to the number of individuals with the same haplotype.

3.3.2. Laboratory Procedures

DNA from fecal, hair and blood samples was extracted and processed following the procedures described in Nater et al. (2011). We used a set of 12 human-derived (Goossens et al. 2005) and 15 species-specific microsatellite markers (Nietlisbach et al. 2010) to genotype the orangutan samples. In order to minimize genotyping errors due to allelic drop-out, we followed the real-time PCR approach from Morin et al. (2001). PCR conditions and fragment length analysis are described in Arora et al. (2010) and Nietlisbach et al. (2010). We were able to genotype 112 out of 162 samples for at least 24 microsatellite loci. The identity check revealed three and two samples that were present as a triplicate and a duplicate, respectively, resulting in 109 unique genotypes.

For the sequencing of the hypervariable region I (HVRI) of the mtDNA d-loop, we used the same primers, PCR conditions and sequencing chemistry as Arora et al. (2010), resulting in a final alignment of 457 base pairs. Some sequences were from samples with insufficient DNA quantity for successful microsatellite genotyping. To avoid duplicates in the HVRI dataset, we only included sequences from individuals that had either a distinct genotype or were sampled more than 50 kilometers apart from other samples in the dataset, resulting in 123 HVRI sequences. The sequences are deposited on GenBank under the accession numbers JQ962945–JQ962972.

3.3.3. HVRI Median-joining Network

A median-joining network (Bandelt et al. 1999) using all HVRI sequences was drawn using NETWORK v4.6.0.0 and NETWORK PUBLISHER v1.3.0.0 (http://www.fluxus-engineering.com). An epsilon value of zero and equal weighting of all nucleotide positions was used for the network presented here. Using higher epsilon values or differently weighted transitions/transversions did not change the basic structure of the network.

3.3.4. Summary Statistics

We computed summary statistics and genetic differentiation measures for HVRI sequences and autosomal microsatellites using ARLEQUIN version 3.5.1.2 (Excoffier & Lischer 2010). For both mitochondrial and autosomal datasets, we incorporated only samples with reliable provenance information. Based on this information, we divided the sample set *a priori* into seven sampling regions (Table 3.1).

Table 3.1: Summary statistics for all examined orangutan sampling regions

		HVRI				Autosomal microsatellites				
Sampling region	Habitat[a]	$N_{Samples}$	θ_π[b]	HD[c]	N_e[d]	$N_{Samples}$	H_E[e]	θ_H[f]	N_e[g]	Census[h]
Tripa (TR)[i]	PSF	7	12.78	0.95	6,808	9	0.64	1.68	4,197	~380
North Aceh (NA)	DF	10	0.79	0.51	389	10	0.61	1.60	3,990	~350
West Leuser (WL)	PSF	28	3.78	0.54	2,013	21	0.61	1.61	4,023	~3000
Central Leuser (CL)	DF	14	0.44	0.27	237	15	0.59	1.56	3,901	~1100
Langkat (LK)	DF	26	1.40	0.80	747	24	0.64	1.66	4,162	~1050
Batu Ardan (BA)	DF	8	0.78	0.46	417	9	0.59	1.57	3,929	~300
Batang Toru (BT)	DF	18	0.96	0.65	503	8	0.63	1.63	4,069	~550

[a], Prevailing habitat type; PSF, peat-swamp forest; DF, dry-land forest (Husson et al. 2009); [b], estimate of $\theta=N_e\mu$ based on the mean pairwise corrected nucleotide distance; [c], haplotypic diversity (Nei 1987); [d], effective population size, based on a mutation rate of 1.643×10^{-7} per site per year and a generation time of 25 years; [e], mean expected heterozygosity; [f], estimate of $\theta=4N_e\mu$ based on the mean expected heterozygosity; [g], effective population size, based on a mutation rate of 10^{-4} per locus per generation; [h] , estimated census size (Wich et al. 2008); [i], The sampling region of Tripa includes coastal and highland areas.

To assess pairwise population differentiation, we calculated the differentiation measures Φ_{ST} (HVRI, Excoffier et al. 1992) and R_{ST} (microsatellites, Slatkin 1995). We used the Tamura & Nei distance correction (Tamura & Nei 1993) with a gamma value of 0.219 for the calculation of the genetic distance matrix for Φ_{ST}, as determined by the model selection test with jMODELTEST version 0.1.1 (Posada 2008).

To infer the long-term effective population size N_e of the seven sampling regions, we calculated the estimators θ_π (based on the mean pairwise genetic distance between sequences; Tajima 1983) and θ_H (based on the heterozygosity of microsatellites; Ohta & Kimura 1973). Additionally, we used a likelihood-based estimator of θ (referred to as θ_L) using the software LAMARC v2.1.6 (Kuhner 2006). We applied the GTR+I nucleotide substitution model (Lanave et al. 1984) for the HVRI sequence data, which is the best-fitting of the supported models inferred by jMODELTEST, and the stepwise mutation model for the microsatellite data. The analysis was performed for each sampling region separately, and we used the Bayesian sampler with two chains of 1,000,000 steps each, sampling every 20^{th} step and discarding the first 5,000 samples as burn-in. The prior distribution of θ ranged from 10^{-5} to 10 (uniform on a natural logarithmic scale) and the starting value of θ was set to 0.01.

The different estimators of θ were used to calculate N_e, with θ equaling $N_e\mu$ for mitochondrial and $4N_e\mu$ for autosomal markers. Thus, these estimators allow inferring N_e from a single population sample if the mutation rate is known. We used a mutation rate of 4.108×10^{-6} per site per generation for HVRI (Soares et al. 2009), assuming a generation time of 25 years (Wich et al. 2009a), or 1×10^{-4} per locus per generation for the autosomal microsatellites (Schlötterer 2000).

3.3.5. Autosomal Genetic Structure

To assess genetic structure based on autosomal microsatellites, we first performed a principal component analysis (PCA) using the covariance-standardized method as implemented in the software GENALEX v6.41. Next, we used the Bayesian clustering algorithm implemented in the software STRUCTURE version 2.3.3 (Pritchard et al. 2000) to identify distinct genetic clusters in the dataset. Because both methods do not require making *a priori* assumptions about genetic structure, we were able to include samples with unknown provenance. For the STRUCTURE analysis, we used the admixture model with correlated allele frequencies, a burn-in length of 3×10^5 steps followed by 3×10^6 MCMC steps. We ran the analysis with K values ranging from 1 to 10. For each K we performed 10 independent runs and averaged the ln Pr(Data|K) statistic over all iterations. Since the Pr(Data|K) estimator has been shown to overestimate K, as it frequently plateaus at higher values than the true number of K (Evanno et al. 2005), we also calculated the delta K statistic (Evanno et al. 2005), which gives a conservative estimate of K.

3.3.6. Migrant Detection

To assess the level of subpopulation connectivity, we identified individuals in the dataset that were either direct migrants or first generation offspring of direct migrants and local individuals. To achieve this, we used two different methods. First, given the strong geographic clustering of mtDNA haplotypes (Nater et al. 2011), we checked the median-joining network for individuals with reliable provenance that clustered with samples from

another geographic region in order to detect direct migrants. Second, we used a Bayesian approach to assign individual genotypes to different subpopulations as either local individuals, direct migrants or F_1 admixed individuals, as implemented in the software BAYESASS 1.3 (Wilson & Rannala 2003). For this, we pre-assigned the individuals to the three different clusters identified in the previous STRUCTURE analysis and ran the MCMC analysis two times independently for 2.4×10^7 steps each, including a burn-in of 4×10^6 steps, with sampling every 2,000 steps. Both runs combined resulted in a total of 20,000 assignments for each individual.

3.4. Results

3.4.1. HVRI Median-joining Network

The median-joining network (Figure 3.1B) showed a strong structuring of mtDNA haplotypes into four geographically distinct clusters: 1) Batang Toru, 2) Langkat, 3) Tripa, West Leuser, Central Leuser and Batu Ardan (referred to as West Alas cluster), and 4) North Aceh. We did not observe any haplotype sharing among these four clusters in our dataset of individuals with reliable provenance information.

3.4.2. Summary Statistics

The division of mitochondrial haplotypes into four distinct clusters as apparent in the mtDNA network correlated well with the Φ_{ST} statistic of genetic differentiation, as all comparisons between different clusters were highly significant (Table 3.2, above diagonal). However, within the West Alas cluster, the sampling region of Tripa was also significantly differentiated from all other regions in the same cluster. This differentiation points to highly different haplotype frequencies between Tripa and the other regions within this cluster, as these all share haplotypes among each other.

The R_{ST} measures for the microsatellites revealed additional information about the population structure beyond female philopatric patterns (Table 3.2, below diagonal). Three main patterns emerged. First, Batang Toru, the only sampling region south of Lake Toba, was highly differentiated from all other regions. Second, in contrast to high mtDNA differentiation, Tripa showed low R_{ST}-values to most other sampling regions, except Batang Toru. Third, the region of Langkat showed low differentiation to North Aceh, Tripa and Batu Ardan.

Table 3.2: Pairwise population differentiation values for HVRI (Φ_{ST}, above diagonal) and autosomal microsatellites (R_{ST}, below diagonal)

Φ_{ST}/R_{ST}	TR	NA	WL	CL	LK	BA	BT
TR	-	0.89*	0.58*	0.70*	0.95*	0.61*	0.97*
NA	0.05*	-	0.94*	0.99*	0.98*	0.98*	0.99*
WL	0.02	0.06*	-	0.04	0.96*	0.01	0.98*
CL	0.04*	0.11*	0.02	-	0.99*	0.02	1.00*
LK	0.02	0.02	0.05*	0.05*	-	0.98*	0.99*
BA	0.05	0.07*	0.07*	0.08*	0.00	-	0.99*
BT	0.12*	0.17*	0.14*	0.10*	0.08*	0.12*	-

* $p<0.05$

The different estimators of θ revealed consistent patterns among the seven sampling regions, but estimates of θ for the microsatellite loci were consistently higher for θ_L as compared to θ_H (Supporting Table S3.2). We found that the genetic diversity estimates based on mtDNA and the corresponding N_e varied extensively across the different sampling regions (Table 3.1), as expected from the large differences in density estimates and habitat areas (Wich et al. 2008; Husson et al. 2009). In general, the estimated effective population sizes were similar to the census size estimates for most sampling regions (Wich et al. 2008). There was one striking exception. Tripa on the northwest coast exhibited high sequence diversity and a N_e of nearly 7,000. This contrasts with the estimated census size of less than 400 individuals. The Tripa region also showed a positive Tajima's D statistic and a multimodal pairwise mismatch distribution of HVRI sequences, indicating a recent population decline, while most other regions exhibited negative values of D and unimodal mismatch distributions, indicating recent expansions (Supporting Table S3.2 and Supporting Figure S3.3). In contrast to the large regional variability for mtDNA, autosomal estimates of genetic diversity and N_e were remarkably similar among sampling regions (Table 3.1).

3.4.3. Autosomal Genetic Structure

The PCA revealed a geographically defined structure in the autosomal microsatellite data (Figure 3.2). The first principal component (PC) explained 25.11% of the total variance and distinguished between the sampling regions west and east of the Alas River. The region south of Lake Toba, Batang Toru, clusters with the regions east of the Alas River and cannot be distinguished with the first PC only. The second PC, explaining a further 18.07% of the variance, separated Batang Toru from all sampling regions north of Lake Toba. Therefore, by combining both PCs (explaining 43.18% of the total variance), there appears to be three clusters of sampling regions, separated from each other by the Alas River and Lake Toba. The separation was, however, not perfect, as the regions of WL, TR, BA and CL showed outliers within the variation of other regions. The additional PCs did not seem to contain any further information about geographic structuring of genotypes (Supporting Figure S3.4).

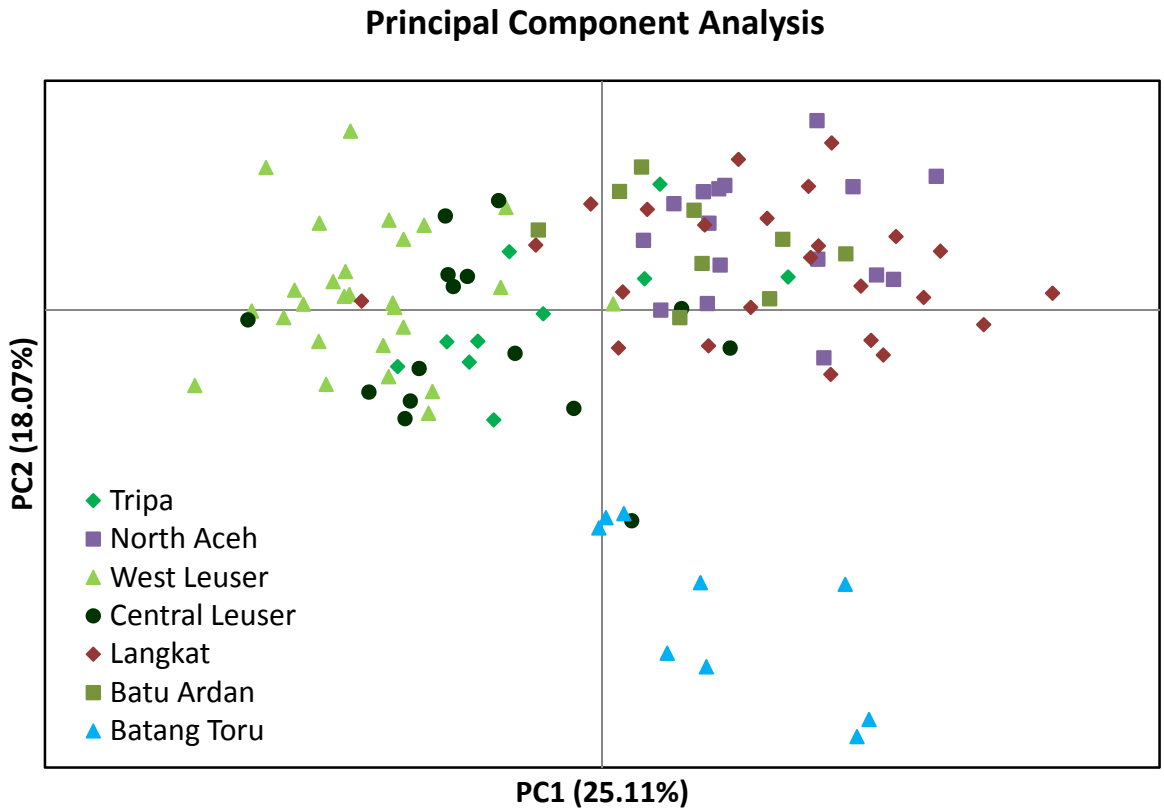


Figure 3.2: Principal component analysis of the autosomal microsatellite markers for all seven sampling regions

The STRUCTURE analysis resulted in a clear signal for substructure in the Sumatran autosomal microsatellite dataset. Highest delta K was achieved for three clusters, while Pr(Data|K) peaked at five clusters (Figure 3.3). At K=3, the clusters corresponded largely to the mtDNA haplotype clusters described above, with some exceptions (Figure 3.4). First, the North Aceh and Langkat regions grouped together. Second, the region of Batu Ardan, which in the HVRI network assigned to the West Alas cluster, showed for autosomal markers a clear affinity to the Langkat and North Aceh regions. Third, the separation between the two genetic clusters north of Lake Toba (West Alas and Langkat/North Aceh) was not as sharp as for the mtDNA, as regions close to the geographic boundaries of the two clusters revealed a number of individuals with admixed genotypes. In contrast, samples from south of Lake Toba (Batang Toru) showed much less signals of admixture. Patterns of genetic admixture were also evident when the membership coefficients Q for each cluster were plotted in ranked order for all individuals for each cluster (Figure 3.5). While all three curves showed two asymptotes at Q=0 and Q=1, multiple samples had Q-values between 0.2 and 0.8 (13 for West Alas, 13 for Langkat/North Aceh, and one for Batang Toru), indicating admixed ancestry.

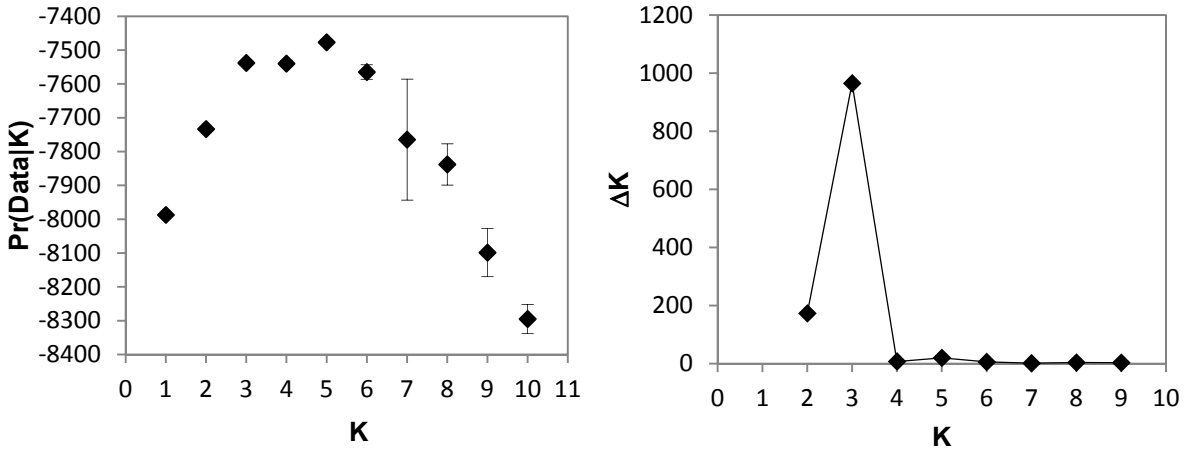


Figure 3.3: Inference of the number of cluster K in the initial STRUCTURE runs. For each number of clusters K, we ran 10 iterations with the same settings. (A) Mean probability of the data given K clusters. The error bars represent ±1 standard deviation. The highest probability is achieved with K=5. (B) Delta K statistic, showing a clear mode at K=3.

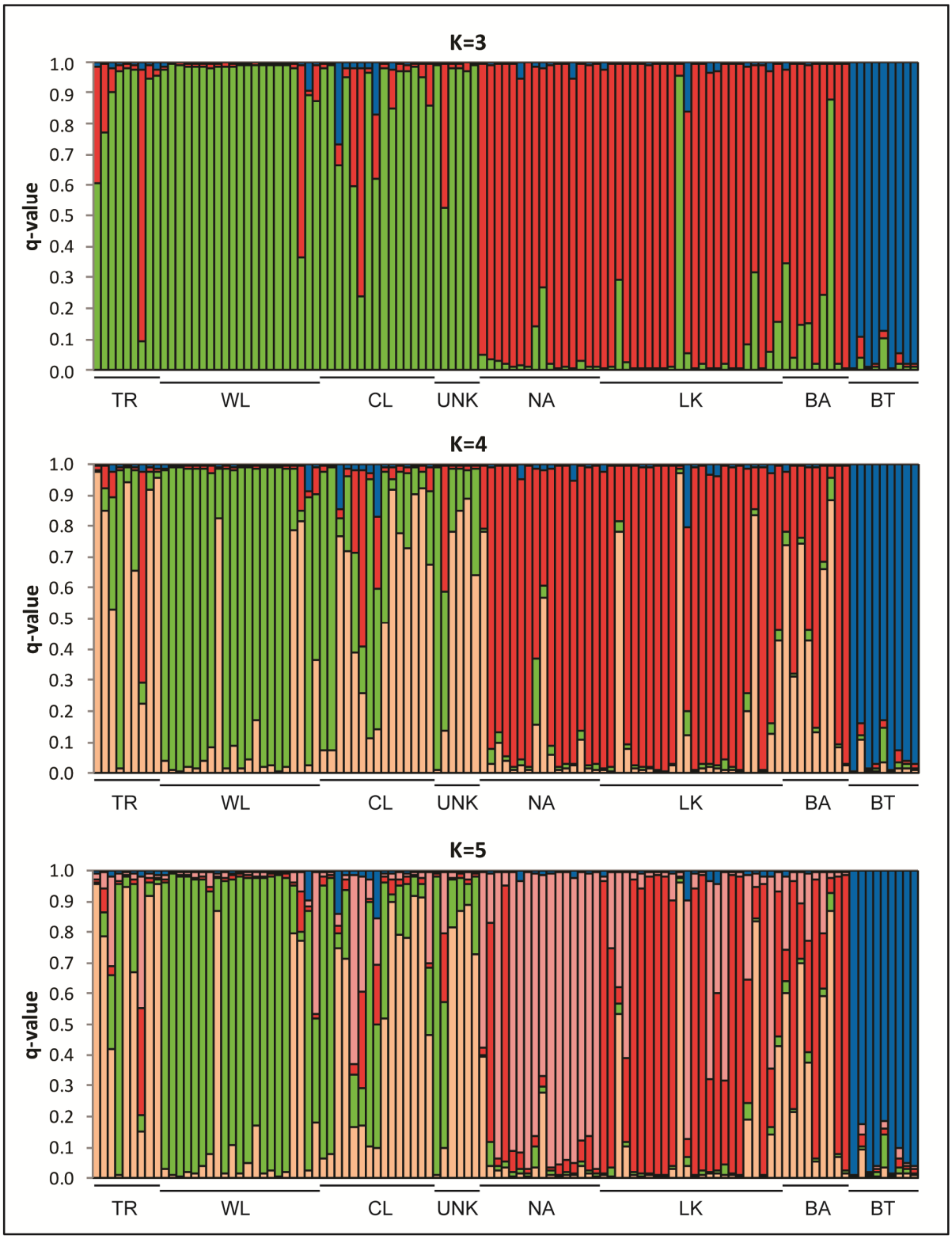


Figure 3.4: Results of the STRUCTURE analysis for three different numbers of clusters (K). The membership coefficients Q shown are for the run with the highest likelihood for each K. The most probable number of cluster according to the delta K statistic is K=3. Samples are grouped by sampling region. The assignment is based on provenance record and mtDNA haplotype. UNK refers to samples with unknown provenance and ambiguous mtDNA assignment (belonging to the West Alas cluster).

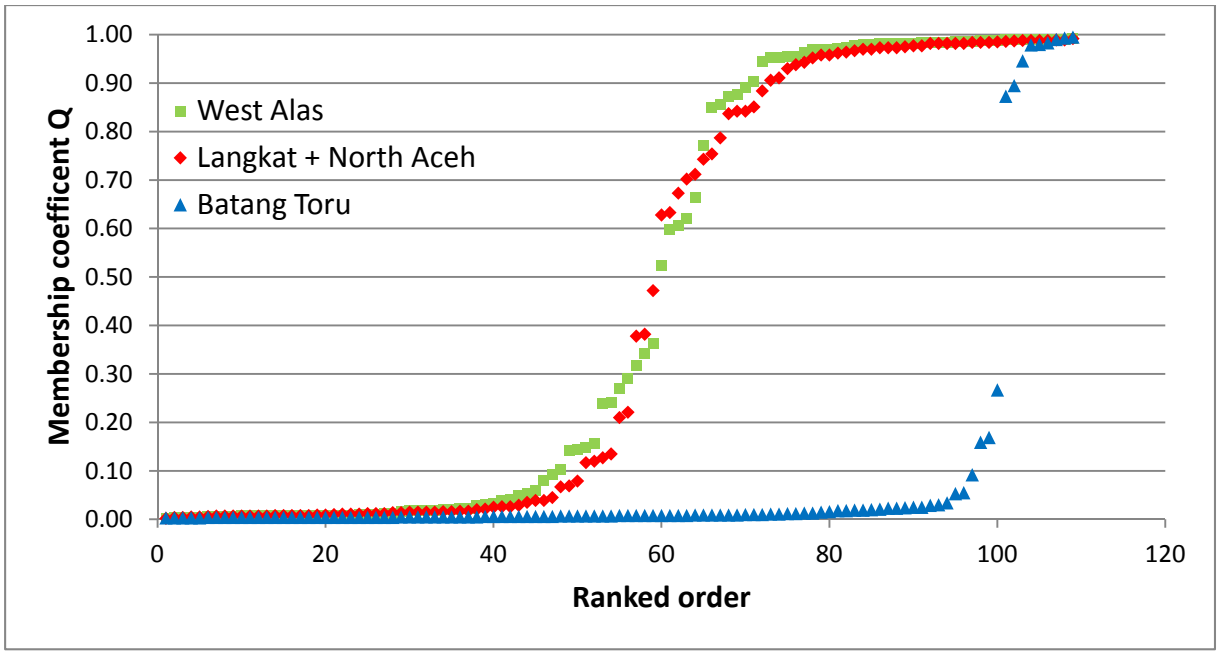


Figure 3.5: Membership coefficients Q plotted in ranked order for each of the three clusters inferred in the STRUCTURE analysis with K=3.

A higher number of K did not result in a better resolution of sampling regions (Figure 3.4). Since STRUCTURE often only identifies the uppermost level of hierarchical genetic structure (Evanno et al. 2005), we repeated the analysis for each of the three geographically defined clusters separately, using only samples that showed a membership coefficient of higher than 0.6 for a certain cluster in the first STRUCTURE analysis. None of the three clusters showed any sign of further substructure, as K=1 returned the highest Pr(Data|K) values for all three clusters.

To test if part of the partitioning of the mitochondrial or autosomal genetic diversity can be explained by habitat type, we performed an AMOVA analysis with ARLEQUIN, where we divided the dataset into two groups corresponding to habitat type (peat-swamp forest versus dry-land forest, see Table 3.1). We included only samples from the West Alas cluster, as this is the only autosomal cluster that contains both habitat types. For autosomal microsatellites, habitat differences explain only 0.22% of the total variance, while over 97% is found within sampling regions (Table 3.3). For the mtDNA diversity, the variance component between habitat types is negative, indicating complete absence of any partitioning of genetic variance between habitat types.

Table 3.3: AMOVA of mitochondrial and autosomal microsatellite data between peat-swamp and dry-land forests within the West Alas cluster

	mtDNA		Autosomal microsatellites	
	Variance	% Variance	Variance	% Variance
Between habitat types	-0.74	-24.55	0.19	0.22
Among sampling regions, within habitat types	1.94*	64.71	1.87	2.17
Within sampling regions	1.80*	59.83	84.34*	97.61

* $p<0.05$

3.4.4. Migrant Identification

All individuals showed congruence between their provenance record and their assigned mtDNA cluster. We did, however, identify three females and two males with high Q-values (>0.6) for a cluster that did not match their mtDNA haplotypes and provenance (K=3, Figure 3.4). These individuals are unlikely to be direct migrants from the autosomal cluster they were assigned to in the STRUCTURE analysis. Rather, their natal range is indicated by their mtDNA haplotype, given that female orangutans have been shown to exhibit strong philopatric tendencies.

The BAYESASS analysis assigned migrant status to three of the five individuals previously identified in the STRUCTURE analysis as admixed or assigned to a cluster that did not match their mtDNA haplotype. In total, we found five individuals which have a less than 50% probability of being local in the cluster defined by their mtDNA (Table 3.4). Only in one case, however, could we identify an admixed individual with significant statistical support ($p<0.05$ of being local). This individual was a female with reliable provenance information, originating from the upper Alas valley in the Langkat region, and carrying an mtDNA haplotype from the Langkat cluster. Her genotype, however, had a high membership coefficient to the West Alas cluster (Q=0.955).

Table 3.4: List of individuals that show a probability of less than 0.5 to originate from the sampling cluster.

Sample Number	Sampling region[a]	Sex	mtDNA[b]	Q-value[c]	BAYESASS[d] Local	Direct migrant	Admixed
BA2	BA (LK+NA)	Female	WA	0.876 (WA)	0.088 (LK+NA)	0.359 (WA)	0.553 (WA)
LK3	LK (LK+NA)	Female	LK	0.702 (LK+NA)	0.494 (LK+NA)	0.010 (WA)	0.496 (WA)
LK27	LK (LK+NA)	Female	LK	0.955 (WA)	0.004 (LK+NA)	0.365 (WA)	0.632 (WA)
LK7	LK (LK+NA)	Male	LK	0.673 (LK+NA)	0.409 (LK+NA)	0.002 (WA)	0.589 (WA)
TR4	TR (WA)	Male	WA	0.884 (LK+NA)	0.443 (WA)	0.228 (LK+NA)	0.329 (LK+NA)

[a], autosomal genetic cluster to which most of the samples from the listed sampling regions assign is written in parentheses: WA, West Alas cluster, LK+NA, Langkat/North Aceh cluster, BT, Batang Toru cluster; [b], mtDNA cluster assignment; [c], highest Q-value in the STRUCTURE analysis with K=3; [d], Posterior probabilities of the three classes in the BAYESASS analysis.

3.5. Discussion

Our study is the first to precisely locate and describe the geographic structuring of genetic diversity on mitochondrial and autosomal levels across the whole range of Sumatran orangutans. We were able to quantify the genetic diversity present within each of the seven sampling regions by analyzing the highly polymorphic HVRI region of the mtDNA and used that information to infer long-term effective population sizes of each sampling region. These estimates correlate strongly with recent census size estimates for most regions (Wich et al. 2008). Not surprisingly, the highest effective population sizes were observed for peat-swamp forests on the west coast of northern Sumatra, which also have the highest population density estimates (Husson et al. 2009). In one region, however, N_e and census size were in stark contrast to each other: the area of Tripa showed an extraordinary high mitochondrial HVRI diversity and corresponding N_e in a comparatively small geographic region, which contains only an estimated 380 individuals. This signal points to a massive recent decline in the subpopulation size, which might have been caused by the dramatic and on-going habitat degradation in this area (van Schaik et al. 2001; Gaveau et al. 2009). It is plausible to assume that the lowland area along the northwest coast of Aceh was once completely covered with continuous peat-swamp forest and harbored thousands of orangutans (Gaveau et al. 2009). After decades of deforestation, current estimates indicate that all forests in the Tripa region will be irrecoverably lost by 2015/16 if forest destruction/conversion will continue at its current rate (Tata et al. 2010; Wich et al. 2011). There are other prominent examples in the literature highlighting discrepancies between large long-term N_e and small census sizes, which are linked to anthropogenic pressures. For example, heavy exploitation of gray (*Eschrichtius robustus*) and humpback whale (*Megaptera novaeangliae*) stocks due to whaling has led to dramatic population declines not reflected by long-term N_e (Roman & Palumbi 2003; Alter et al. 2007).

In contrast to the varying HVRI diversity found within different regions across the island, we obtained very homogenous genetic diversity estimates among sampling regions for autosomal microsatellite markers, resulting in N_e estimates of around 4,000 or 10,000 individuals for each of the seven regions, depending on the estimator of θ. This striking discrepancy as compared to the HVRI estimates is most likely caused by pronounced male-biased dispersal and strong female philopatric tendencies in orangutans (Galdikas 1995; Singleton & van Schaik 2002; Morrogh-Bernard et al. 2011; Arora et al. 2012; Nietlisbach et al. 2012; van Noordwijk et al. 2012). Field studies have shown that female orangutans preferentially establish their home range overlapping with the home ranges of their maternal kin (Singleton & van Schaik 2002; van Noordwijk et al. 2012). Thus, mitochondrial DNA does get hardly, if at all, exchanged among neighboring geographic regions, and mtDNA diversity well reflects the number of orangutans in the different local subpopulations. Males, in contrast, leave their natal area, a pattern linked to inbreeding avoidance (Pusey & Wolf 1996). Intense male-male competition (Utami Atmoko et al. 2009b) may force young males to cover large distances before being able to settle down (Nietlisbach et al. 2012). Such widely dispersing males might distribute newly arisen alleles in the whole meta-population and recover alleles that have been lost locally due to genetic drift, thereby homogenizing the allele frequencies of autosomal

markers among sampling regions. Thus, the highly similar levels of autosomal diversity in contrast to the large differences in mtDNA diversity across the island are a clear indicator of considerable male-mediated gene flow among these regions. The panmictic distribution of Y-haplotypes on Sumatra (Nater et al. 2011) provides further evidence for this male-driven homogenization of the gene pool.

Due to the use of multiple independent autosomal markers, we were able to investigate male-mediated gene flow in more detail. The cluster analysis with STRUCTURE showed that the strength of male-driven gene flow is not sufficient to completely homogenize allele frequencies among sampling regions, thus resulting in a clear pattern of geographically structured autosomal variation. The three clusters identified in the autosomal dataset were defined by geographical features. It appears that eruptions of the Toba volcano (Chesner et al. 1991) isolated the orangutans from Batang Toru, the region south of it, from the rest of the species occurring north of it. The high pairwise R_{ST}-values across Lake Toba provide further evidence of strong separating effects of the Toba eruptions, which have also led to a deep divergence of mtDNA haplotypes north and south of the caldera (Nater et al. 2011). The forests between these two areas might have been connected between major eruptions, but the combination of periodic separation and strong female philopatry has served to keep the populations from homogenizing. North of Lake Toba, the Alas River, part of the Barisan graben running the length of Sumatra (Verstappen 1973), divides the remaining regions into two distinct genetic clusters. The Alas valley was likely repeatedly blocked by volcanic material from the nearby Toba eruptions, turning the upper Alas river into a large lake for prolonged periods (van Schaik & Mirmanto 1985). This damming of the Alas River might have promoted the structuring of the gene pool north of Lake Toba. Interestingly, the habitat type does not seem to play a significant role in the structuring of autosomal diversity in Sumatran orangutans, indicating that dispersing males do not prefer to migrate to areas that ecologically resemble their natal habitat, and thus prevent more fine-tuned adaptation of orangutans to local habitat types.

Even though the STRUCTURE analysis revealed strong geographical structuring of the autosomal gene pool, we nevertheless found clear signals for recent gene flow across the island. First, the two sample regions of Langkat and North Aceh cannot be distinguished in the STRUCTURE analysis, even though these regions show a mitochondrial divergence of 0.85 Ma (Nater et al. 2011). Therefore, the observed low autosomal differentiation (R_{ST}=0.02) points towards considerable levels of male-mediated gene flow after the two subpopulations were separated from each other. If this migratory contact with the Langkat region can be maintained, it will greatly help reducing inbreeding pressure on the small North Aceh subpopulation. As a second signal of gene flow, we found many admixed individuals in the STRUCTURE plot (Figure 3.4). Interestingly, these individuals were mostly sampled in regions close to the boundary of autosomal clusters, like Tripa, Central Leuser and Langkat, supporting the idea of recent gene flow. Third, we were able to identify multiple individuals with substantial likelihoods of having paternal ancestry from another cluster. While only one individual shows good statistical support for being admixed ($p<0.05$), it should be kept in mind that we sampled only an estimated 0.7 to 4.6% of all individuals per sampling region. Moreover, we only investigated migration among major autosomal clusters and not individual sampling regions, due to the impossibility to reliably discriminate them genetically.

Further investigation of the provenance of admixed individuals hinted toward an important corridor for gene flow between genetic clusters. Three of the five individuals identified as having admixed ancestry originate from the upper Alas valley near Blangkejeren, while a fourth admixed individual has been confiscated in the highlands of the Tripa area. These locations are all close to the area where the supposed boundaries of the West Alas, North Aceh and Langkat clusters meet, and this highland area contains orangutan habitat with resident subpopulations. The presence of clear migration signals in this area underlines its critical importance as a connection among major subpopulations of Sumatran orangutans and therefore deserves special habitat conservation efforts.

Special consideration also needs to be given to the region of Batu Ardan, where there is a clear discrepancy between autosomal data and mtDNA structure, possibly due to male-mediated migration. This region, located between the Alas River and Lake Toba, shows a strong affinity of mtDNA haplotypes to the West Alas cluster, even though it is located on the opposite (eastern) side of the major Alas River. In fact, Batu Ardan shares a common haplotype with all regions on the western side, but also has two derived haplotypes that do not occur elsewhere. This supports the notion that the small Batu Ardan subpopulation could be the result of a recent colonization event from the western side of the Alas, probably due to a loop cut-off of the meandering river (Nater et al. 2011). However, for autosomal markers, we found that Batu Ardan reveals a high affinity to the adjacent Langkat/North Aceh cluster, from which it is separated by a deep river valley. This river might be passable by orangutans near its headwaters, allowing males to bring in autosomal alleles from the Langkat region. The notion that the recolonization from the west side of the Alas and subsequent influx of males from Langkat occurred after the forests recovered from the devastating Toba supereruption around 73 kya (Chesner et al. 1991) is tempting but cannot yet be proven with the data at hand.

Sumatran orangutans are genetically deeply structured into at least three autosomally distinct clusters, despite regular male-mediated gene flow between the West Alas and the Langkat/North Aceh clusters, which occurred at least up to very recently and is probably still on-going. However, continuing habitat degradation is threatening the existence of orangutans on Sumatra in two ways. First, due to the shrinkage of suitable habitat area, the local subpopulation census sizes will be further reduced. Already today, only one of the three autosomal clusters, West Alas, harbors well over 1,000 individuals. Second, through the destruction of important corridors for migration, genetic exchange with neighboring subpopulations will be disrupted. Both effects combined will inevitably lead to a substantial loss of genetic diversity with all its negative consequences (Reed & Frankham 2003). Especially the only remaining subpopulation south of Lake Toba, Batang Toru, is highly threatened in this regard. Given the genetic uniqueness of the orangutans in this area on both the mitochondrial and autosomal level and the fact that most of the forest in this area has no protected status (Wich et al. 2011), urgent measures are needed to preserve this indispensable reservoir of genetic diversity of Sumatran orangutans.

Orangutans are the least gregarious and the most arboreal of all great apes (Delgado & Van Schaik 2000). As such, comparing the observed patterns in Sumatran orangutans with those of other great ape species will aid the inference of factors underlying the observed population

structure in these taxa. Previous genetic studies on great apes showed that rivers are one of the most important factors in shaping population structure and subspecies boundaries (e.g. bonobos: Eriksson et al. 2004; Goossens et al. 2005; gorillas: Anthony et al. 2007; chimpanzees: Becquet et al. 2007; Bornean orangutans: Arora et al. 2010). Our study supports these findings by identifying the Alas River as a major division line of genetic diversity within the range of Sumatran orangutans. Moreover, volcanic activities of the Toba region during the last 1.2 million years (Chesner et al. 1991) played another major role in the structuring of genetic diversity in Sumatran orangutans. Such a pattern of long-lasting isolation caused by volcanic activities has so far not been documented for great apes.

Given that Sumatran orangutans are critically endangered, knowledge of the extent to which human-induced habitat degradation is affecting the population structure is of critical importance for conservation efforts. Bergl and Vigilant (2007) revealed a pronounced substructure in the small Cross River gorilla population (*Gorilla gorilla diehli*) largely following the patterns of forest connectivity. Likewise, Goossens et al. (2005) showed that in Bornean orangutans, subpopulations in many of the isolated forest lots on the same side of the Kinabatangan River in Sabah, Malaysia, are significantly differentiated from each other, despite their close geographic proximity. Both studies highlight the adverse effects of anthropogenic forest degradation on the dispersal abilities of forest dwelling primates. Interestingly, we did not observe similar signals in Sumatran orangutans, despite their strict arboreality and the heavy forest exploitation within their range (Rijksen & Meijaard 1999). The Sumatran subpopulations appear to be more effectively connected through male dispersal for two reasons. First, the uninhabited mountain regions connecting subpopulations are forested, and thus dispersing males, who have been sighted at altitudes of up to 2000 m above sea level (Rijksen 1978), can move through them. Second, Sumatran forests provide suitable habitat to higher altitudes than Bornean ones due to the *Massenerhebung* effect (van Schaik et al. 1995), and this makes it easier for migrating males to cross rivers at their headwaters.

The example of the Sumatran orangutan demonstrates that even species with a geographically very limited range can show strong underlying genetic structure, caused by geographical barriers, habitat discontinuities, limited dispersal, and long population persistence. Correspondingly, genetic diversity might be mainly found among local subpopulations rather than within, and local extinctions carry a serious risk of losing a substantial part of a species' total genetic diversity. Our study highlights the need to assess the genetic make-up of endangered species in detail, identify local subpopulation boundaries, and focus conservation efforts on maintaining dispersal corridors among genetic clusters.

3.6. Acknowledgements

We are very grateful to Ellen Meulman, Andrea Permana, Izumi, Zulfikar, Tony Weingrill, Helga Peters, Gregoire Bertagnolio, Gail Campbell-Smith, Yenny Saraswati and Rachmad Wahyudi for collecting samples. Corinne Ackermann and Kai Ansmann performed valuable laboratory work for this study. We thank Erik Willems for contributing the map of northern Sumatra. Furthermore, we highly appreciate the support of the following institutions: Indonesian State Ministry of Research and Technology (RISTEK), Indonesian Institute of Sciences (LIPI), Sabah Wildlife Department, Taman National Gunung Leuser (TNGL),

Borneo Orangutan Survival Foundation (BOSF), Leuser International Foundation (LIF), and Badan Pengelola Kawasan Ekosistem Leuser (BPKEL). This work was financially supported by the Swiss National Science Foundation (grant number 3100A-116848 to MK and CvS); the Messerli Foundation; the A.H.-Schultz Foundation; and the Claraz Schenkung.

Chapter 4:

Reconstructing the Demographic History of Orangutans using Approximate Bayesian Computation

Alexander Nater[1], Maja P. Greminger[1], Natasha Arora[1], Carel P. van Schaik[1], and Michael Krützen[1]

(1) Anthropological Institute & Museum, University of Zurich, Zurich, Switzerland

Keywords: *Pongo* spp., demographic history, Approximate Bayesian Computation, Population structure

Author contributions:

Alexander Nater conceived the study, conducted genetic sampling on Sumatra, developed and performed laboratory procedures, conducted statistical analyses, and wrote the manuscript.

Maja P. Greminger conceived the study, performed laboratory procedures and edited the manuscript.

Carel P. van Schaik and Michael Krützen conceived the study and edited the manuscript.

Natasha Arora performed laboratory procedures.

4.1. Abstract

Investigating how different evolutionary forces have shaped patterns of DNA variation within and among extant species requires detailed knowledge of the demographic history, as these patterns are the result of demographic, selective and random processes. Orangutans, whose distribution is currently restricted to the Southeast Asian islands of Borneo (*Pongo pygmaeus*) and Sumatra (*Pongo abelii*), have likely experienced a complex demographic history, influenced by recurrent climate changes, volcanic activity and anthropogenic pressures. Previous studies have tested simplified demographic models, often using a small number of genetic samples with unknown geographic origin. Thus, it remains unknown to what extent unrepresentative population sampling, population substructure and oversimplified models have led to misleading conclusions. Approximate Bayesian Computation (ABC) allows testing complex demographic models using different types of genetic markers in a combined analysis. In our ABC approach, we employed the most extensive genetic data set of wild orangutans to date, including autosomal and Y-chromosomal microsatellite genotypes, as well as autosomal, X-chromosomal and mitochondrial sequence data. We tested the fit of 8 different demographic models, which we designed based on current knowledge of orangutan genetics and behavioral ecology. We found that a complex demographic model, incorporating population substructure within Bornean and Sumatran orangutans, a recent bottleneck on Borneo, as well as a recent population decline on Sumatra, best explains the currently observed patterns of genetic variation in orangutans. Based on this demographic model, we estimate that the two orangutan species diverged ~900 kya, with subsequent heavily male-biased migration between the two islands until ~107 kya. We confirm the previous finding that Sumatran orangutans exhibit a deep split of populations north and south of Lake Toba, probably caused by multiple eruptions of the Toba volcano. In addition, we found signals for a strong decline in all Sumatran populations 9–15 kya, probably associated with hunting by human colonizers. In contrast, Bornean orangutans experienced a severe bottleneck ~61 kya, followed by a population expansion and substructuring starting ~22 kya, which we link to an expansion from a common refugium during the last glacial period. Thus, we show that orangutans, like other non-human great apes, went through drastic changes in population size and connectedness, caused by the recurrent contraction and expansion of rainforest habitat during Pleistocene glaciations, and probably also by the impact of hunting by early humans. Our results contrast with a previous study, which fitted simplified demographic models to genomic data, resulting in a species divergence ~400 kya, followed by a continuous exponential growth on Sumatra and a continuous decline on Borneo. The discrepancies between these finding and our results demonstrate that caution has to be exerted when using oversimplified demographic models and potentially inappropriate sampling schemes to reconstruct demographic history.

4.2. Introduction

Reconstructing the processes of how evolutionary forces have shaped patterns of DNA variation in different species has for long been one of the most important goals of evolutionary biology. These patterns are the result of both adaptive and non-adaptive processes, and the debate about the relative importance of natural selection and random genetic drift in shaping genetic diversity within and among species is still ongoing (Hahn 2008; Wagner 2008; Nei et al. 2010). A common approach to detect signals of selection aims at identifying genomic regions that show marked deviations in DNA variation patterns from the expectations under a neutral equilibrium model (reviewed in Nielsen 2005). However, under certain demographic scenarios, such as population size changes or population subdivision, random drift can result in similar deviations from a neutral equilibrium model as selection (e.g. Tajima 1989; Andolfatto & Przeworski 2000; Hahn et al. 2002; Nielsen et al. 2005; Teshima et al. 2006; Excoffier et al. 2009). Thus, confounding effects of demographic processes can only be disentangled from selective signals if the demographic history is explicitly taken into account when formulating the expectations under the neutral model against which observed patterns of DNA variation are tested (Wall et al. 2002; Haddrill et al. 2005; Nielsen et al. 2005; Stajich & Hahn 2005). Consequently, methods to reconstruct the demographic history of natural populations have recently evoked large interest among evolutionary geneticists, as recent technical advances allow conducting genome-wide studies of selection in a large variety of species (reviewed in Metzker 2010).

Orangutans, currently restricted to two distinct species on Borneo (*Pongo pygmaeus*) and northern Sumatra (*Pongo abelii*) (Wich et al. 2008), are the only Asian great apes and phylogenetically most distant to humans (Groves 2001). Their ancestral position in the lineage leading to African great apes and modern humans has evoked great interest in this taxon in the overall effort to reconstruct the adaptive evolutionary history of great apes in general and humans in particular (Locke et al. 2011). However, orangutans might have experienced a complex demographic history, as their distribution has been subject to major changes during the Pleistocene. The ancestors of extant orangutans have sequentially colonized the islands of the Sunda archipelago arriving from the Southeast Asian mainland (von Koenigswald 1982a; Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). Since then, their population history was strongly influenced by geological and climatic events: Rising and falling sea levels cyclically connected and isolated the islands of Sundaland, allowing for potential migration between the islands at certain points in time (Voris 2000). Major volcanic eruptions, mainly on Sumatra and Java, might have led to the extinction of local orangutan populations and subsequent recolonizations (Muir et al. 2000). Especially the Toba supereruption, which occurred ~73 kya on Northern Sumatra and is considered to be the most powerful volcanic eruption within the last 25 million years (Chesner et al. 1991), is thought to have had severe consequences for the flora and fauna on Sundaland (Williams et al. 2009). In the Late Pleistocene, all populations on the mainland, southern Sumatra and Java went extinct (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). While climatic changes during the Pleistocene might be responsible for the southward shift of the distribution and the disappearance of orangutans on the mainland (Jablonski 1998), anthropogenic factors, mainly

prehistoric hunting by hunter-gatherer societies, are likely to have played a significant role in the decline and extinction of orangutans populations on insular Southeast Asia (Delgado & Van Schaik 2000).

Genetic signals of these past demographic changes have been found in studies of genetic diversity in extant orangutan populations on Borneo and Sumatra. Most genetic studies analyzing autosomal and mitochondrial data agree that Sumatran orangutans show a higher level of sequence diversity and corresponding effective population size (N_e) (i.e. Muir et al. 2000; Zhang et al. 2001; Steiper 2006; Locke et al. 2011), even though Sumatran orangutans have a much smaller current census size and a more restricted distribution than Borneans (~6,600 vs. ~54,000 individuals, Wich et al. 2008). Muir et al. (2000) and Steiper (2006) explained this discrepancy by the possibility that the extant populations on northern Sumatra were once connected by regular gene flow with large populations on southern Sumatra, the mainland, and Java. The Bornean populations, on the other hand, were more isolated and might have exchanged migrants with other populations only sporadically during the glacial periods (Voris 2000; Warren et al. 2001; Steiper 2006). Interestingly, the Y-chromosomal diversity in orangutans shows an opposite pattern as compared to mtDNA and autosomal data, with a smaller N_e on Sumatra than Borneo (Nater et al. 2011). Since orangutans show heavily male-biased dispersal (Morrogh-Bernard et al. 2011; Arora et al. 2012; Nietlisbach et al. 2012; van Noordwijk et al. 2012), the recent coalescence time of Y chromosomes on Sumatra (~4.2 kya) suggests that the extant populations on Sumatra did not experience any recent genetic exchange with other differentiated populations, as such gene flow would have brought distinct Y-chromosomal haplotypes into the populations on Sumatra.

Different studies tried to reconstruct the genetic relationships between the two orangutan species by using molecular clock methods to date the population split between Bornean and Sumatran orangutans. These efforts resulted in a broad spectrum of divergence estimates, ranging from 1.1 Ma (Warren et al. 2001) to 1.3–7.7 Ma (Steiper 2006), depending on the choice of genetic markers and the calibration method. While most estimates were based solely on mitochondrial DNA (mtDNA) data and therefore represented only the history of maternal lineages, the study by Steiper (2006) included sequence data from both mitochondrial and nuclear loci. This approach gives a better representation of genome wide genetic divergence, and showed that the two orangutan species have been genetically isolated since the Early Pleistocene. However, by combining mtDNA sequences with Y-chromosomal loci, and therefore investigating both maternal and paternal population histories, Nater et al. (2011) found the divergence of male specific lineages between Borneo and Sumatra to be much more recent as compared to mtDNA (~170 kya vs. ~2.1 Ma), indicating that long-range dispersal of males led to genetic exchange between the two orangutan species during glacial periods.

More recently, Locke et al. (2011) used extensive single-nucleotide polymorphism (SNP) data from whole genome resequencing of five Bornean and five Sumatran orangutans to model the demographic history of the two species. They found that a model with a population split ~400 kya with subsequent gene flow between Borneo and Sumatra fits the observed data best. Furthermore, Locke and colleagues inferred that Sumatran orangutans underwent a continuous exponential population growth since the population split, while Bornean orangutans were subject to a continuous exponential decline.

Given the enormous amount of genetic data involved, the study by Locke et al. (2011) has been widely seen as the most accurate reconstruction of demographic history in orangutans to date. However, the demographic modeling approach by Locke and colleagues did not take several idiosyncrasies of the orangutan biology into account, thus severely limiting the conclusions that could be drawn from their findings. First, the study incorporated only data from five captive individuals each from Borneo and Sumatra without further provenance information. This limited genetic sampling is unlikely to represent the entire genetic diversity present on both islands. Moreover, given this lack of detailed sample provenance, the modeling was restricted to models that treated Bornean and Sumatran orangutans each as a single panmictic population. Previous studies, however, have shown that both Bornean and Sumatran orangutans are genetically deeply structured (Warren et al. 2001; Arora et al. 2010; Nater et al. 2011), and especially Sumatran orangutans exhibit high genetic differentiation of populations north and south of Lake Toba (Nater et al. 2011; Nater et al. 2013). Ignoring such population substructure often produces misleading results regarding population size changes (Stadler et al. 2009; Chikhi et al. 2010; Peter et al. 2010).

Second, Locke et al. (Locke et al. 2011) did not further investigate temporal patterns of gene flow between Borneo and Sumatra. Thus, it remains unknown if gene flow between the two islands occurred until the end of the last glacial period ~10 kya, when land bridges between Borneo and Sumatra were severed (Voris 2000). Last, Locke and colleagues did not test complex demographic models including population bottlenecks or recent declines, as suggested in previous genetic studies. For example, genetic signals of a bottleneck with subsequent population expansion on Borneo might be linked to a glacial refugium or the impact of the Toba supereruption ~73 kya (Steiper 2006; Arora et al. 2010), and patterns of a recent population decline in Sabah, Borneo are most likely attributable to recent anthropogenic pressures (Goossens et al. 2006a).

Reconstructing the demographic history of a species has long been hindered by the fact that full-likelihood methods were restricted to relatively simple demographic models (e.g. Wilson et al. 2003; Hey & Nielsen 2004; Kuhner 2006), which might not capture all relevant processes in complex demographic settings. This restriction is mainly caused by the fact that the computation of the likelihood function of complex demographic models with many parameters is either impossible or computationally too intensive, especially for large data sets (Marjoram et al. 2003). Approximate Bayesian Computation (ABC) allows circumventing these problems by approximating the likelihood functions with simulations of genetic data under a given demographic model (Beaumont et al. 2002; Marjoram et al. 2003). In order to estimate the model parameters, parameter values are drawn from predefined prior distributions and used to simulate genetic data that is matching the observed data in the type of markers and number of loci. Both observed and simulated data are then reduced to a set of summary statistics and the Euclidian distance between the observed and the simulated summary statistics is calculated. Based on the subset of simulations with the smallest Euclidian distance between observed and simulated data, the posterior distribution of the model parameters can be approximated.

Here we present an ABC modeling approach of the demographic history of orangutan based on autosomal and sex-linked marker systems. We aim to improve the current knowledge of

demographic history by applying three major improvements over previous studies. First, we capitalize on the knowledge base of behavioral ecology and population genetics of orangutans in order to test realistic demographic models. Second, due to our extensive set of orangutan samples with detailed and reliable provenance, we are able to investigate models incorporating population substructure in both orangutan species, which allows us to disentangle changes in population size from confounding effects due to changes in population structure. Third, by combining autosomal and sex-linked markers into a combined demographic analysis, we make use of the specific information content of different marker systems in species with heavily sex-biased dispersal. Due to strong female philopatry in orangutans (Galdikas 1995; Arora et al. 2012; van Noordwijk et al. 2012), mitochondrial markers contain information about population split times without confounding influence of gene flow. In contrast, Y-chromosomal loci should have more power than autosomal markers to reveal low levels of male-mediated gene flow. Additionally, the inclusion of sex-specific markers will allow us to investigate patterns of sex-biased dispersal over evolutionary significant time spans and large distances.

4.3. Materials & Methods

4.3.1. Sample Collection

Our sample set for this study included orangutan samples used in previous genetic studies of orangutans (Arora et al. 2010; Nater et al. 2011; Nater et al. 2013). These samples were either fecal and hair samples non-invasively collected from wild populations or blood samples collected from rehabilitant orangutans. Geographic provenance of samples from rehabilitant orangutans was confirmed based on their mtDNA haplotypes (Figure 4.1), which has been shown to be a reliable indicator for the natal area in orangutans (Arora et al. 2010; Nater et al. 2011). Sample details and DNA extraction procedures are described in the aforementioned studies. The collection and transport of samples was conducted in strict accordance with Indonesian, Malaysian and international regulations. Samples were transferred to Zurich under the Convention on International Trade in Endangered Species (CITES) (permits 09717/IV/SATS-LN/2010, 07279/IV/SATS-LN/2009, 00961/IV/SATS-LN/2007, 06968/IV/SATS-LN/2005, and 4872).

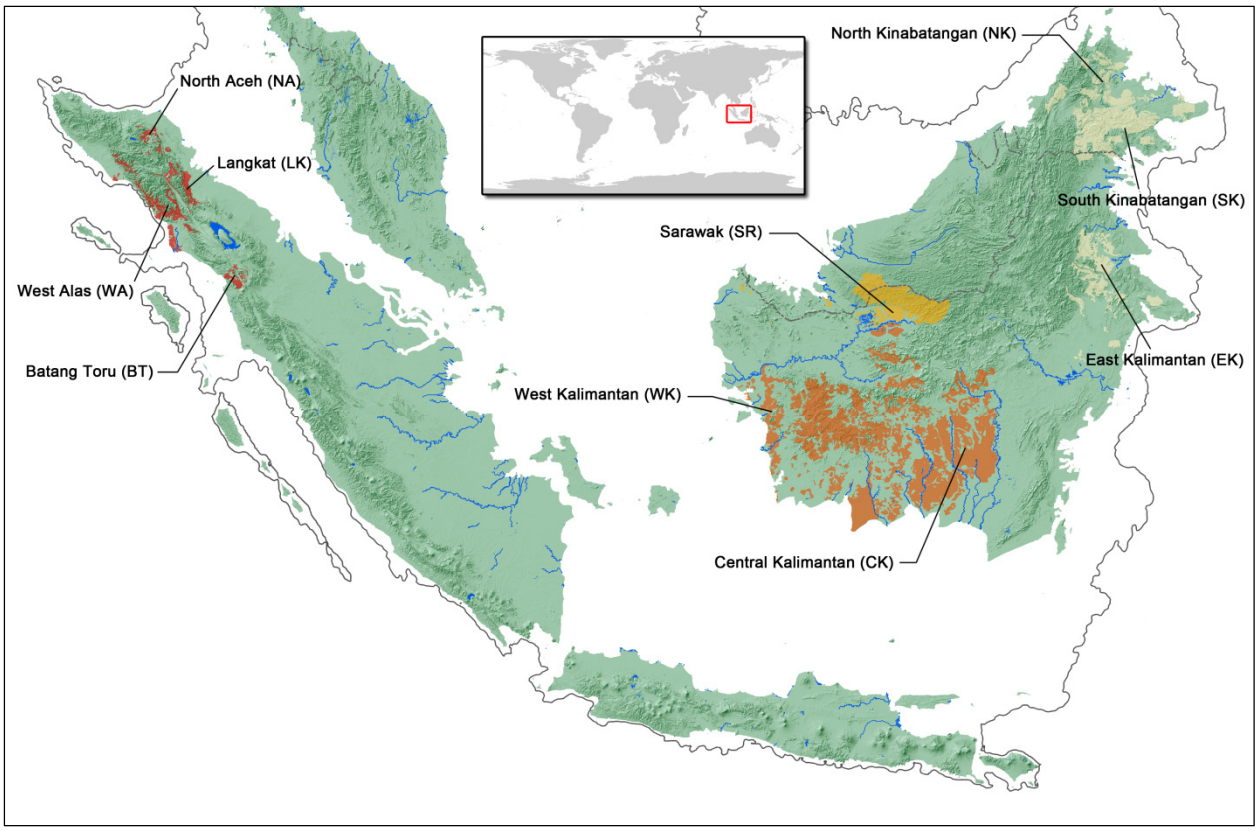


Figure 4.1: Map of sampling regions in Sundaland used for the demographic modeling. The colored shadings represent the current distribution of the Sumatran orangutans and the three subspecies of Bornean orangutans. The grey line indicates the extent of the exposed Sunda shelf during the LGM (19–26 kya, -120 meters below current sea level).

4.3.2. PCR Amplification, Sequencing and Genotyping

We used a genetic data set that incorporated a combination of autosomal, X-chromosomal, mitochondrial and Y-linked loci. The microsatellite data contained genotypes of 25 autosomal microsatellite markers from a total of 237 individuals (Arora et al. 2010; Nater et al. 2013). We also included sequences from three mtDNA genes with a total length of 1,355 bp from 118 individuals (Nater et al. 2011), and Y-chromosomal haplotypes based on 11 Y-linked microsatellite loci from 129 individuals (Nater et al. 2011). We complemented the data set by additionally sequencing 8,055 bp of the non-coding X-chromosomal region Xq13.3 (Kaessmann et al. 2001) in 37 individuals and four non-coding autosomal regions (Fischer et al. 2006) of a total of 8,255 bp in 20 individuals. Sample sizes in each sampling region for the different genetic markers are provided in Table 4.1.

Table 4.1: Sample seizes for different marker systems and ten geographic regions

Sampling region[a]	mtDNA	Y-STRs	Autosomal STRs	Autosomal regions	Xq13.3
North Kinabatangan (NK)	6	10	32	4	3
South Kinabatangan (SK)	13	15	76	2	3
East Kalimantan (EK)	7	9	34	2	5
Sarawak (SR)	8	2	12	2	1
Central Kalimantan (CK)	9	9	68	2	2
West Kalimantan (WK)	9	8	32	4	4
Batang Toru (BT)	8	8	18	4	3
North Aceh (NA)	7	15	32	6	3
Langkat (LK)	14	15	66	10	6
West Alas (WA)	37	38	104	4	7
Total	**118**	**129**	**474**	**40**	**74**

[a], sampling regions corresponding to Figure 4.1. Sample sizes are given as number of sampled chromosomes. The light gray shading refers to Bornean populations, middle gray to Sumatran populations north of Lake Toba, and dark gray to the Sumatran population south of Lake Toba.

The primers used for PCR amplification and sequencing of the X-chromosomal and autosomal regions are described in the Supporting Table S4.1. The PCRs contained 10 ng genomic DNA, 0.16 µl Phire Hot Start DNA Polymerase, 1x Phire Reaction Buffer (both Finnzymes) containing 1.5 mM $MgCl_2$, 0.1 mM dNTPs and 0.1 µM each of forward and reverse primer in 8 µl total volume. PCR amplifications were performed in a Veriti Thermal Cycler (Applied Biosystems) with the following parameters: Initial denaturation at 98°C for 30 seconds, 40 cycles of 98°C for 10 seconds, primer specific annealing temperature for 10 seconds and 72°C for 40 seconds, followed by a final extension step at 72°C for 5 minutes. Cycle sequencing was performed with BigDye Terminator v3.1 chemistry on a 3730 DNA Analyzer (both Applied Biosystems). We used SEQUENCING ANALYSIS v5.3.1 (Applied Biosystems) for raw data analysis. The SEQMAN program of the LASERGENE 8 software package (DNASTAR) was used to trim and align the sequences. We used the program PHASE v.2.1 (Stephens et al. 2001) to infer haplotypes of autosomal and X-chromosomal sequences. Heterozygous positions with phasing probabilities of less than 0.95 were coded with IUPAC ambiguity codes.

4.3.3. Data Visualization

We used the Bayesian clustering algorithm implemented in the software STRUCTURE v2.3.3 (Pritchard et al. 2000) to identify and visualize genetic structure in the autosomal microsatellite data set. We applied the admixture model with correlated allele frequencies, a burn-in length of 3×10^5 steps followed by 3×10^6 Markov chain Monte Carlo (MCMC) steps, running the analysis with the number of clusters K ranging from 1 to 10. We performed ten iterations per K and averaged the likelihood of the data Pr(D|K) over all iterations for each K to calculate the deltaK statistic (Evanno et al. 2005), which we used as a criterion to select the most probable number of clusters in the data set.

We constructed phylogenetic trees for mitochondrial, X-chromosomal and autosomal sequences using the Bayesian MCMC method implemented in BEAST v1.6.2 (Drummond, Rambaut 2007). We applied a TrN+G substitution model (Tamura & Nei 1993) for the mitochondrial alignment and a HKY+G+I model (Hasegawa et al. 1985) for all autosomal and X-chromosomal alignments, as determined by jMODELTEST v0.1.1 (Posada 2008). Each gene trees was rooted with a human and a central chimpanzee sequence from GenBank. The BEAST software was also used to estimate locus-specific mutation rates under a relaxed molecular clock model (Drummond et al. 2006). The calibration of the molecular clock was implemented as described in Nater et al. (2011).

4.3.4. Approximate Bayesian Computation

Model Selection

We attempted to reconstruct the demographic history of orangutans using an ABC approach implemented in the software package ABCtoolbox v1.1 (Wegmann et al. 2010). To achieve this goal, we first performed a model selection procedure, whereby we tested eight different demographic models with increasing levels of complexity (see Supporting Table S4.2 for more details about model parameterization and prior distributions). The tested models can be divided into models assuming a single population each for all Bornean and Sumatran samples, and models incorporating population substructure with six populations on Borneo, one Sumatran population south of Lake Toba, and three Sumatran populations north of Lake Toba (Figure 4.1). We defined this substructure based on the observed patterns of population differentiation for mtDNA markers (Arora et al. 2010; Nater et al. 2011). Since the number of simulated populations differed between these two sets of models, we applied a script that pooled the simulated data into a Bornean and a Sumatran group after each simulation step. Summary statistics were then calculated island-wise rather than population-wise, in order to be able to directly compare models with different levels of complexity.

The first set included four different models, all assuming a single population for each of the two orangutan species (Figure 4.2A). The first model in this set (I2) assumed no migration between the two populations, with the split time as well as Bornean, Sumatran and ancient N_e being estimated. The second model (IM2) incorporated two migration matrices, with asymmetric migration possible after the population split up to a point where migration between Borneo and Sumatra ceased. The third model (IM2-GR) is a further refinement of the second model and additionally allowed the two populations to change size exponentially after the population split. In the fourth and most complex 2-population model (IM2-BN-GR), both populations retained a constant size after the population split, with the possibility for a sudden population size rescale followed by exponential growth, therefore allowing for a bottleneck in both populations.

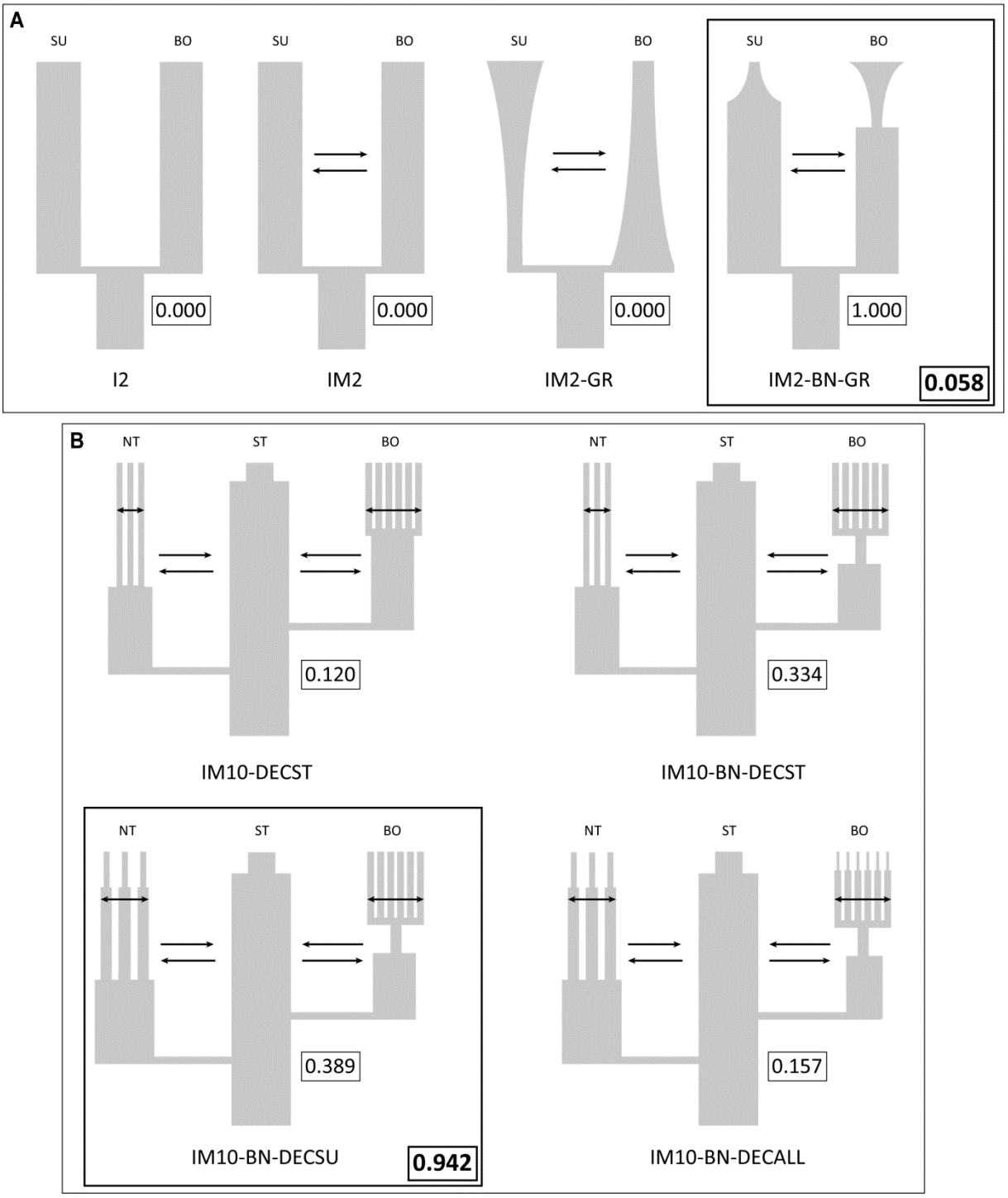


Figure 4.2: Schematic representation of all eight tested demographic models, divided into four 2-population models and four 10-populations models (SU = Sumatra, BO = Borneo, NT = Sumatra north of Lake Toba, ST = Sumatran south of Lake Toba). The posterior probabilities within each group are given next to each model. The best model in each group is marked with a black box and the posterior probabilities when comparing these two best models against each other are given below each box.

The second set included four different models assuming ten populations, as supported by recent work on phylogeographic patterns in orangutans (Arora et al. 2010; Nater et al. 2011) (Figure 4.2B). All models in this set were parameterized with a single parameter for the population sizes on Borneo and north of Lake Toba, respectively, as well as a separate population size parameter for the population south of Lake Toba. We included asymmetric migration rates between Borneo and south of Lake Toba, and between north of Lake Toba and south of Lake Toba. We assumed a symmetric migration rate among all populations within Borneo and among all populations north of Lake Toba. The first 10-population model (IM10-DECST) incorporated a recent population decline south of Lake Toba, based on the historic records of widespread occurrence of orangutans in this area (Rijksen & Meijaard 1999). The second model (IM10-BN-DECST) additionally included a period of reduced population size in the ancestral population on Borneo in order to simulate a refugium. As a further refinement, the third model additionally incorporated a population decline in all Sumatran populations (IM10-BN-DECSU), while the fourth model (IM10-BN-DECALL) included such a decline in all ten populations.

To simulate genetic data under these eight demographic models, we used the software FASTSIMCOAL v1.1.2 (Excoffier & Foll 2011). Simulations for the different marker systems were run with the same set of parameters, whereby the effective population sizes were scaled 1 to 0.75 to 0.25 to 0.25 for autosomal, X-chromosomal, mitochondrial and Y-chromosomal markers, respectively. A parameter representing the proportion of males in the migrant pool (MALEMIG) was defined in order to be able to estimate common migration rates for all marker systems. We then used ARLSUMSTAT v3.5.1.3 (Excoffier & Lischer 2010) to calculate 68 summary statistics for each simulated data set as well as for the observed data set (Supporting Table S4.3). The summary statistics were chosen in order to capture the information in the genetic data about population differentiation, within population diversity, and population size changes. To avoid problems with unreliable phasing, we only used summary statistics that do not require phased sequence data for X-chromosomal and autosomal loci.

We first performed an initial run of 10^6 simulations with the standard rejection sampler (Tavare et al. 1997). These simulations were used for both model selection and validation. To reduce the dimensionality of the summary statistics, we performed a principal component analysis (PCA) in R version 2.12.1 (R Development Core Team 2010) with the standardized summary statistics and extracted the first ten principal components of the simulated as well as the observed data. We applied an ABC-GLM post sampling regression adjustment (Leuenberger & Wegmann 2010) as implemented in ABCtoolbox on the 1,000 simulations with the smallest Euclidean distance to the observed data. To compare different models, we calculated the Bayes factor (BF, Jeffreys 1961) by taking the ratio of the marginal densities of two models as given by ABCtoolbox (Leuenberger & Wegmann 2010). We followed the recommendations given by Jeffreys (1961) to evaluate the support for each of the tested models. Additionally, we calculated the posterior probability of each model defined as the marginal density of a model divided by the sum of marginal densities of all assessed models (Veeramah et al. 2011).

Parameter Estimation

To obtain good estimates of the posterior distributions of the parameters for the selected model, we used a MCMC without likelihood method (Wegmann et al. 2009b). To reduce the dimensionality of the data and extract as much information as possible about the model parameters, we used the first 20,000 simulations with the standard sampler to define the first 15 orthogonal components of the summary statistics that maximize the covariance matrix between summary statistics and model parameters. For this, we applied a partial least-squares (PLS) regression approach (Boulesteix & Strimmer 2007) as implemented in the "pls" R package (Mevik & Wehrens 2007) and using the R script provided in the ABCtoolbox package. This way, a large set of summary statistics is reduced to a number of independent components, whereby summary statistics that are most informative about the model parameters are weighted more than summary statistics that do not show much response to changing parameter values (Wegmann et al. 2009b). The initial simulations were also used to define the tolerance distance based on a tolerance level of 0.1 and to calibrate the transition kernel of the MCMC run with a rangeProp setting of 1 unit of standard deviation (Wegmann et al. 2009b; Wegmann et al. 2010). We then ran a total of 3×10^6 iterations with the MCMC sampler, followed by a ABC-GLM post sampling regression on the 15,000 simulations with the smallest Euclidean distance to the PLS components of the observed summary statistics. Finally, we used R to plot the posterior distributions of important model parameters.

Validation

The performance of ABC in model selection and parameter estimation in complex population genetic settings inevitably suffers from the loss of information when the observed and simulated genetic data are reduced to a set of summary statistics (Robert et al. 2011). This necessitates a careful validation of the employed ABC procedure in order to avoid biases in the approximation of posterior probabilities of evaluated models and the estimation of model parameters. Accordingly, we validated our model selection and parameter estimation approach with three different procedures. First, we investigated the model misclassification rate by generating 100 pseudo-observed data sets randomly drawn from the prior distributions for each of the eight tested models. We then performed the same model selection procedure as with the real observed data and counted the number of assignments to each of the eight models. Second, we assessed the goodness of fit of all tested models to the observed data by calculating the p-value of the observed data under the GLM applied for the post sampling regression using ABCtoolbox (Leuenberger & Wegmann 2010). The p-value is representing the proportion of the retained simulations showing a lower or equal likelihood under the inferred GLM as compared to the observed genetic data (Wegmann et al. 2009a). Third, to increase confidence in the parameter estimates of the selected model, we checked for biased posterior distributions by producing 1,000 pseudo-observed data sets under the selected model with parameter values drawn from the prior distributions. We used ABCtoolbox to calculate the posterior quantiles of the true parameter values within the estimated posterior distributions for each pseudo-observed data set and used a Kolmogorov-Smirnov test for uniformity in R (Wegmann et al. 2009b). Significant deviation from uniformity after sequential Bonferroni correction (Rice 1989) indicate biased posterior distributions (Cook et al. 2006).

4.4. Results

4.4.1. Data Visualization

The STRUCTURE run analyzing the autosomal microsatellite data set resulted in the highest deltaK values for K=2 (Supporting Figure S4.4), clearly separating Bornean and Sumatran individuals (Figure 4.3A). Since STRUCTURE tends to find only the highest level of hierarchical genetic structure in a data set (Evanno et al. 2005), we repeated the analysis separately for each island. This resulted in two and three distinct clusters on Borneo and Sumatra, respectively (Figure 4.3B). The two Bornean clusters separated individuals from south of the Kinabatangan River in Sabah (South Kinabatangan) and East Kalimantan from individuals from Central and West Kalimantan, Sarawak, as well as north of the Kinabatangan River (North Kinabatangan). Further runs incorporating only samples from the same higher-level cluster revealed a total of five distinct genetic clusters within Bornean orangutans, separating nearly all regions except Sarawak, which clusters together with West Kalimantan (Figure 4.3C). In Sumatra, we detected no further hierarchical substructure. Thus, at the lowest level of hierarchal genetic structure, there are a total of eight distinct autosomal clusters (5 on Borneo, 3 on Sumatra) among all sampled orangutans.

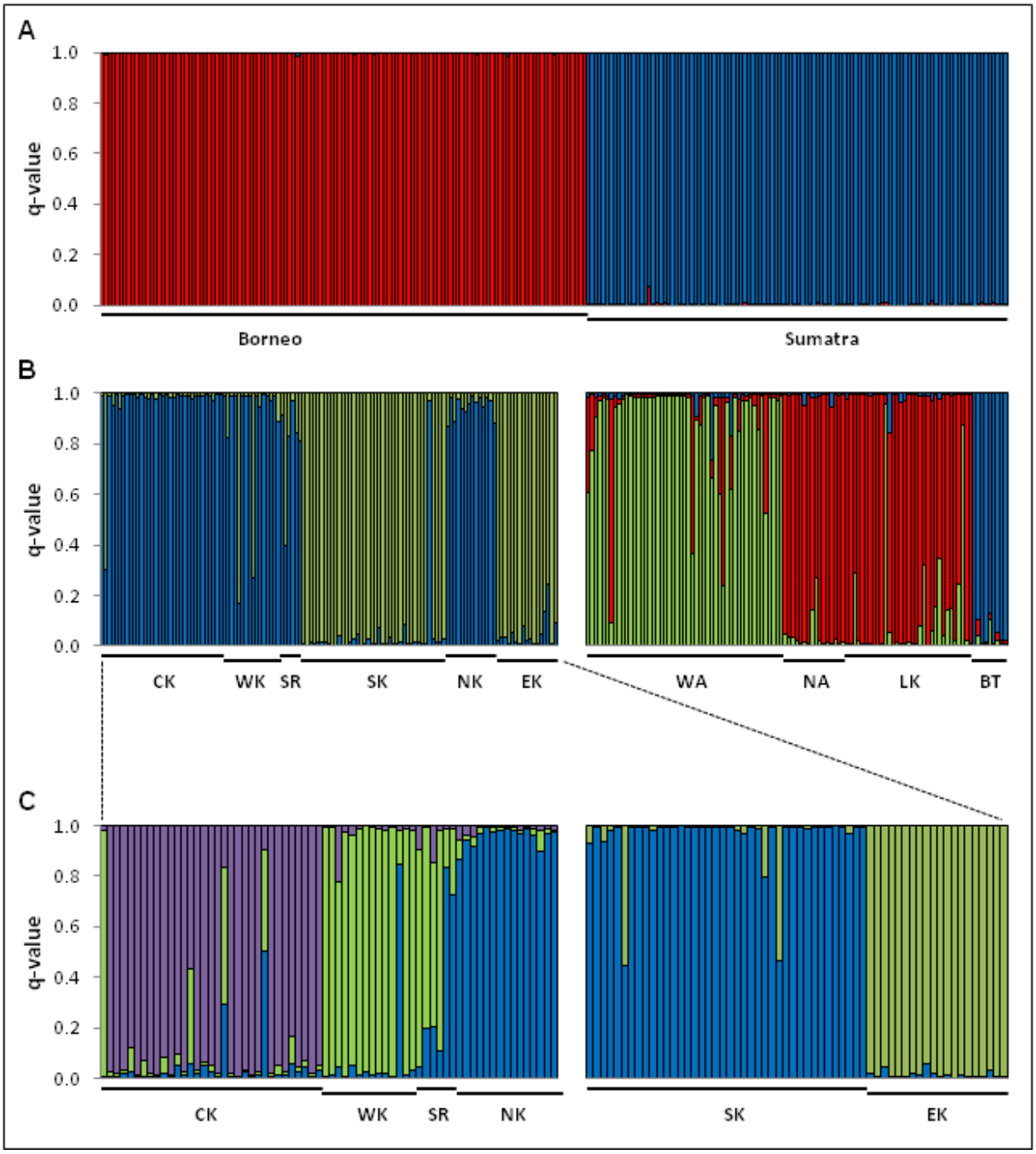


Figure 4.3: Structure plots for 25 microsatellite markers used for the demographic modeling. The three rows of plots correspond to the three levels of hierarchical structure we identified in the complete dataset.

We used a Bayesian MCMC approach to infer gene trees and mutation rates of the autosomal, X-chromosomal and mitochondrial loci, based on our sequence alignments. For the four autosomal loci and the single X-chromosomal locus, the BEAST runs resulted in mean mutation rates of $1.64–3.02\times10^{-8}$ and 1.96×10^{-8} per site per generation, respectively. As expected, the mitochondrial regions showed a mutation rate that was an order of magnitude higher as compared to the nuclear loci (2.80×10^{-7} per site per generation). The phylogenetic trees of the five nuclear loci revealed different topologies compared to the mitochondrial tree

(Figure 4.4). All autosomal regions showed incomplete lineage sorting and in some cases even haplotype sharing between Borneo and Sumatra. For the X-chromosomal region, all Bornean sequences formed a monophyletic group with a comparatively recent common ancestor, while the Sumatran sequences were paraphyletic. The Sumatran population south of Lake Toba, Batang Toru, did not form a distinct cluster for any of the five gene trees.

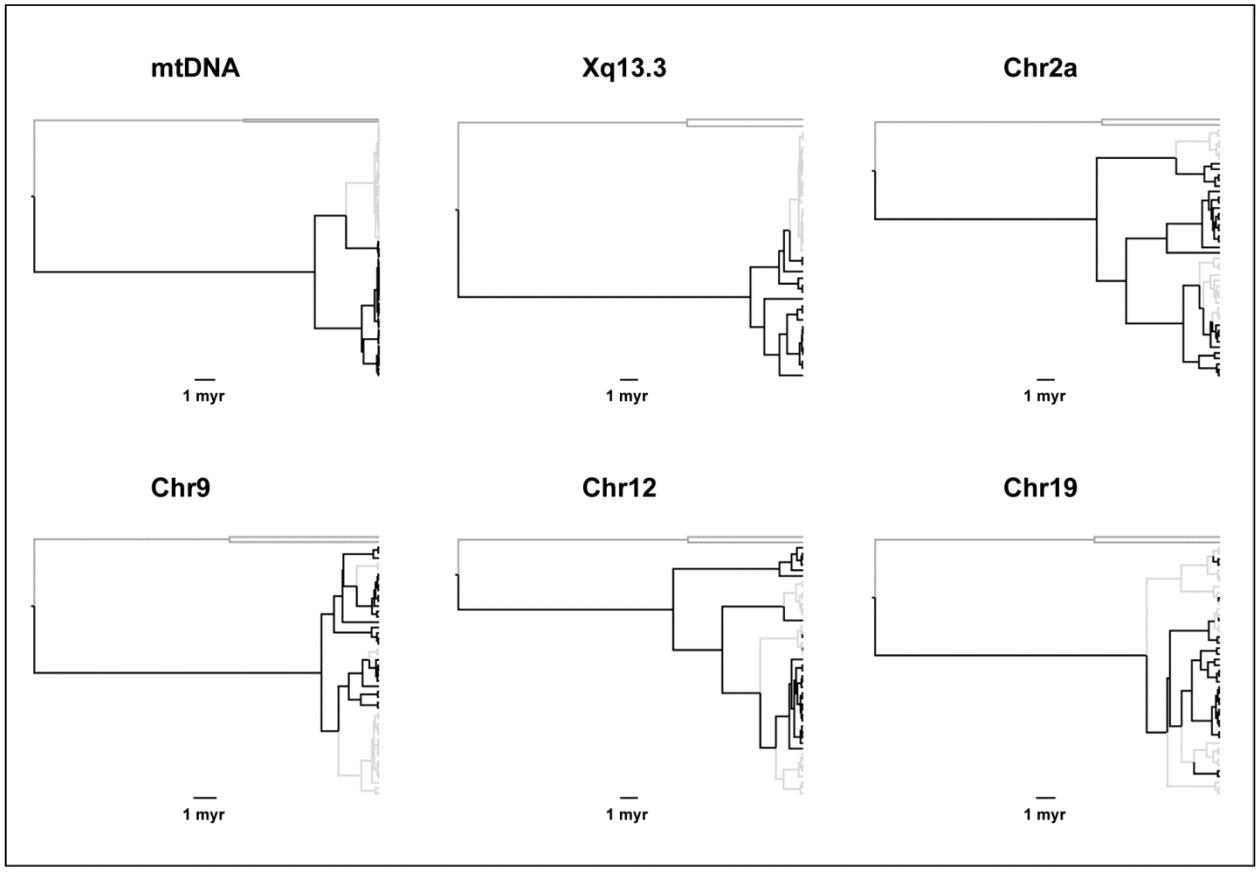


Figure 4.4: Gene trees based on sequence data of six different loci. The tips of black branches refer to Sumatran samples, light gray to Bornean samples, and dark gray to the human and chimpanzee outgroup.

4.4.2. Model Selection

We tested a total of eight different demographic models (Figure 4.2). When comparing the two simplest models that treated Borneo and Sumatra as single populations, but differed in the possibility for migration after the population split (I2 vs. IM2), we found strong support for the model allowing genetic exchange after the split (Log_{10} BF 1.77, Table 4.2). However, such a simple isolation with migration model achieved only a very poor fit to the observed data (GLM p-value 0.005). Of the four 2-population models tested, we observed a decisive support for a model that allowed a sudden change in population size for both populations followed by exponential growth (IM2-BN-GR, Log_{10} BF 3.83). Still, this model did not achieve a good fit to the observed data, as evidenced by a p-value of the observed data under the GLM of only 0.050.

Table 4.2: Log_{10} Bayes factors for all model comparisons

Model	Log_{10} MD[a]	p-value[b]	I2[c]	IM2[d]	IM2-GR[e]	IM2-BN-GR[f]	IM10-DECST[g]	IM10-BN-DECST[h]	IM10-BN-DECSU[i]	IM10-BN-DECALL[j]
I2	-15.95	0.004	-	-1.77	-4.82	-8.65	-9.35	-9.80	-9.87	-9.47
IM2	-14.18	0.005	1.77	-	-3.06	-6.88	-7.59	-8.03	-8.10	-7.71
IM2-GR	-11.13	0.005	4.82	3.06	-	-3.83	-4.53	-4.98	-5.04	-4.65
IM2-BN-GR	-7.30	0.050	8.65	6.88	3.83	-	-0.70	-1.15	-1.21	-0.82
IM10-DECST	-6.60	0.208	9.35	7.59	4.53	0.70	-	-0.45	-0.51	-0.12
IM10-BN-DECST	-6.15	0.388	9.80	8.03	4.98	1.15	0.45	-	-0.07	0.33
IM10-BN-DECSU	-6.08	0.567	9.87	8.10	5.04	1.21	0.51	0.07	-	0.39
IM10-BN-DECALL	-6.48	0.352	9.47	7.71	4.65	0.82	0.12	-0.33	-0.39	-

[a], marginal density of the observed data under the inferred GLM; [b], p-value of the observed data under the inferred GLM; [c], isolation model with two populations; [d], isolation-with-migration model with two populations; [e], isolation-with-migration model with two populations and exponential growth; [f], isolation-with-migration model with two populations and bottleneck followed by exponential growth; [g], isolation-with-migration model with 10 populations and recent decline in population south of Lake Toba; [h], isolation-with-migration model with 10 populations, bottleneck on Borneo and recent decline in population south of Lake Toba; [i], isolation-with-migration model with 10 populations, bottleneck on Borneo and recent decline in all Sumatran populations; [j], isolation-with-migration model with 10 populations, bottleneck on Borneo and recent decline in all populations.

We attributed the poor model fit of all tested 2-population models to population substructure, which differs to a great extent for female- and male-mediated marker systems (Nater et al. 2011). Accordingly, the N_e for each marker system varied to a large degree and could not be described accurately with just one population size parameter per island. In agreement with this notion, we found that the more complex 10-population models achieved a better fit to the observed genetic data better than the 2-population models (Table 4.2). The overall best fitting model (GLM p-value 0.567) was the 10-population model with a refugium in Borneo and a recent decline on Sumatra (IM10-BN-DECSU, Figure 4.5), for which we found strong support over the best 2-population model (IM2-BN-GR, Log_{10} BF 1.21).

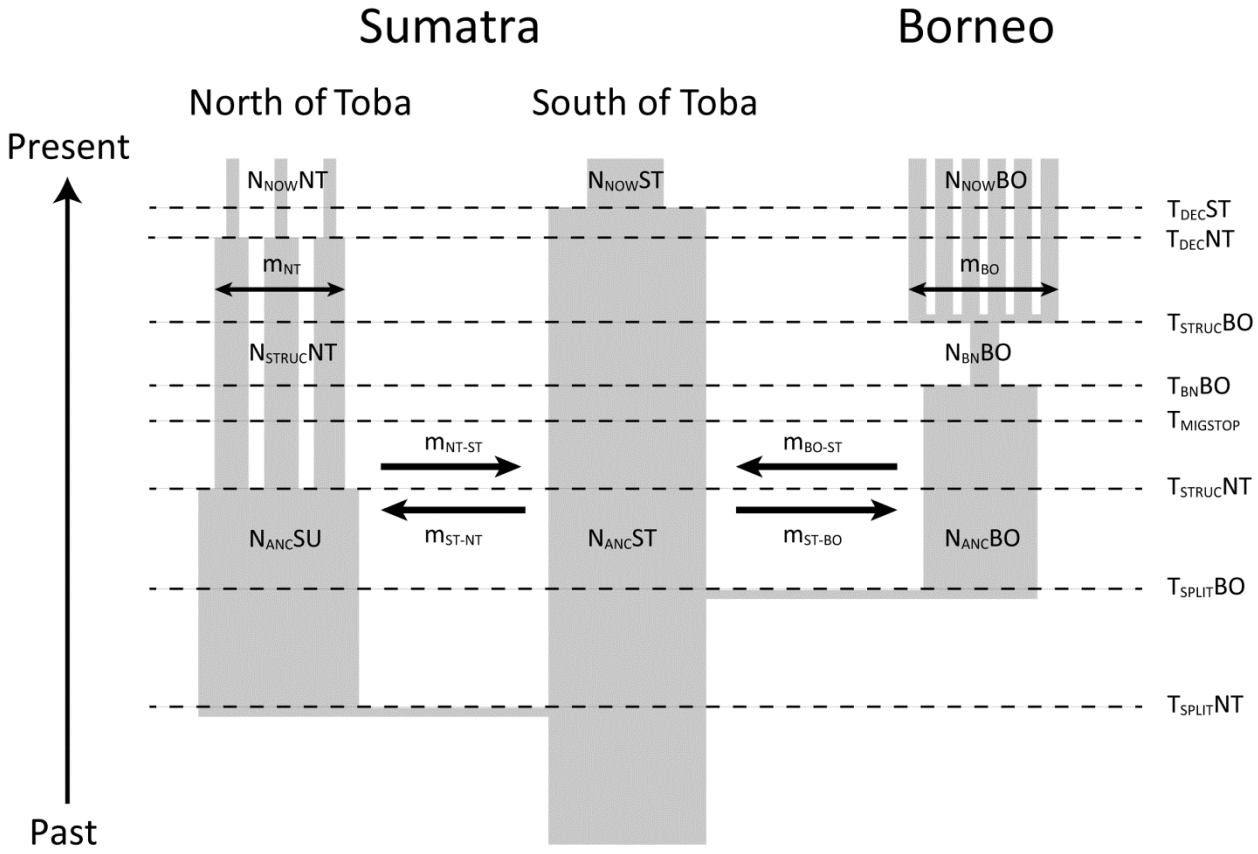


Figure 4.5: Schematic representation of the selected 10-population model with a bottleneck on Borneo and recent population declines in all Sumatran populations (IM10-BN-DECSU).

4.4.3. Parameter Estimation

We estimated the model parameters for the selected 10-population model based on 3×10^6 simulations of the likelihood-free MCMC run (Table 4.3, Figure 4.6). The parameter estimates point to a current N_e of ~880 diploid individuals in each of the six Bornean populations. We found support for a bottleneck on Borneo starting ~61 kya and ending ~22 kya, during which N_e on Borneo was reduced around ten times from an ancestral N_e of ~6,800 individuals to ~650 individuals. The bottleneck on Borneo was followed by population recovery and substructuring, with a current total N_e of all Bornean populations of ~5,300 individuals.

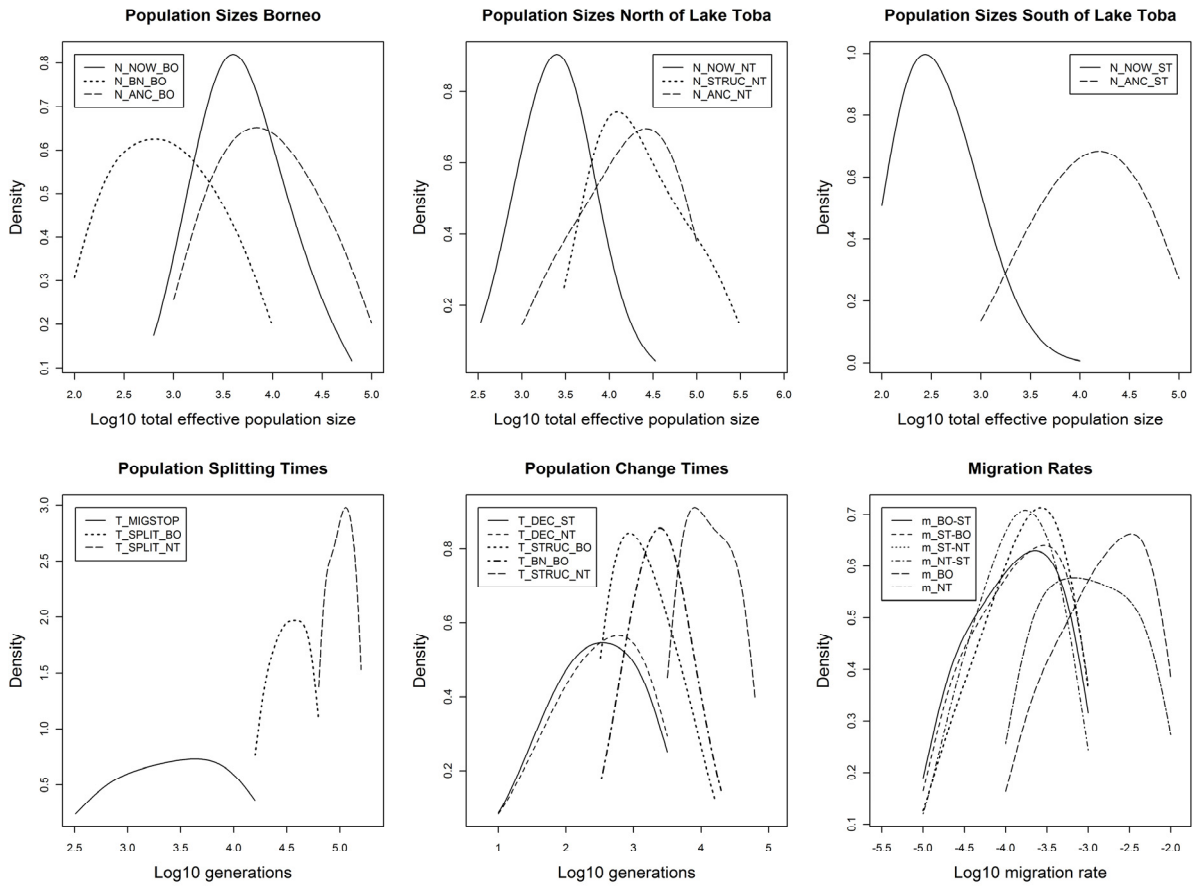


Figure 4.6: Posterior distributions of important model parameters under the selected 10-population model. The abbreviations of the model parameters correspond to the labels in Figure 4.5. For better comparability, the effective populations sizes of the structured meta-populations on Borneo and north of Lake Toba are given as the total effective sizes according to the formula $N_e=D\times N\times(1+(1/(4\times N\times(D-1)\times m)))$, with D corresponding to the number of subpopulations, N to the mean subpopulation size and (D-1)×m to the total migration rate per individual per generation within the meta-population (Nichols et al. 2001).

On Sumatra, the three populations north of Lake Toba suffered a decline ~15 kya from a mean N_e of ~4,100 to currently only ~360 individuals, corresponding to a total N_e in the meta-population north of Lake Toba of ~12,700 and ~1,600 individuals before and after the decline, respectively. We estimated that population structure north of Lake Toba was established ~200 kya, with an ancestral effective population size of ~26,000 individuals. The population south of Lake Toba also went through a decline ~9 kya from a N_e of ~15,700 individuals in the ancestral population to currently only ~280 individuals.

We inferred the population split time between Borneo and south Toba as ~940 kya, and between north and south of Lake Toba as ~2.8 Ma. There is no evidence for asymmetric migration between Borneo and south of Lake Toba, and between south of Lake Toba and north of Lake Toba. Gene flow between Borneo and Sumatra was generally low (0.9–1.8 migrants per generation in each direction) and appears to have ceased ~107 kya. Populations north and south of the Toba caldera on Sumatra historically exchanged ~3 migrants per generation in each direction, but this gene flow dropped to just 0.06–0.16 migrants following the population declines on Sumatra. Migration rates among the populations on Borneo and

north of Toba, respectively, were considerably higher than between the two islands or between populations north and south of Lake Toba. Bornean populations exchanged ~1.6 migrants per generation among each other and populations north of Lake Toba historically exchanged ~5.4 migrants per generation, which then dropped to just ~0.3 migrants. Overall, we obtained a strong signal for highly male-biased migration among populations, with a proportion of males in the migrant pool of 0.87.

Table 4.3: Estimates of the model parameters for the selected 10-population model with a bottleneck on Borneo and a recent decline on Sumatra.

Parameter[a]	R^{2} [b]	Prior	Mode	Mean	90%-HPD[c]
$Log(N_{NOW}BO)$	0.696	unif[2.0,4.0]	2.94 (880)	2.99 (966)	2.44, 3.54
$Log(N_{NOW}NT)$	0.439	unif[2.0,4.0]	2.56 (365)	2.70 (506)	2.01, 3.27
$Log(N_{NOW}ST)$	0.345	unif[2.0,4.0]	2.44 (277)	2.62 (420)	2.01, 3.17
$Log(N_{BN}BO)$	0.030	unif[2.0,4.0]	2.81 (652)	2.94 (868)	2.07, 3.72
$Log(N_{ANC}BO)$	0.099	unif[3.0,5.0]	3.83 (6,826)	3.96 (9,032)	3.12, 4.76
$Log(N_{STRUC}NT)$	0.180	unif[3.0,5.0]	3.61 (4,103)	3.90 (7,889)	3.06, 4.68
$Log(N_{ANC}NT)$	0.064	unif[3.0,5.0]	4.42 (26,127)	4.14 (13,836)	3.39, 4.99
$Log(N_{ANC}ST)$	0.123	unif[3.0,5.0]	4.20 (15,703)	4.08 (12,041)	3.33, 4.92
$Log(T_{STRUC}BO)$	0.115	unif[2.5,4.2]	2.94 (21,931)	3.21 (40,172)	2.51, 3.82
$Log(T_{BN}BO)$	0.090	unif[2.5,4.2]	3.39 (61,027)	3.39 (61,087)	2.70, 4.06
$Log(T_{SPLIT}BO)$	0.009	unif[4.2,4.8]	4.57 (937,972)	4.52 (818,502)	4.27, 4.78
$Log(T_{DEC}NT)$	0.091	unif[1.0,3.5]	2.77 (14,767)	2.45 (7,019)	1.56, 3.49
$Log(T_{STRUC}NT)$	0.007	unif[3.5,4.8]	3.91 (201,210)	4.13 (338,384)	3.58, 4.68
$Log(T_{SPLIT}NT)$	0.002	unif[4.8,5.2]	5.05 (2,828,590)	5.01 (2,534,195)	4.84, 5.18
$Log(T_{DEC}ST)$	0.114	unif[1.0,3.5]	2.55 (8,773)	2.40 (6,228)	1.50, 3.45
$Log(T_{MIGSTOP})$	0.117	unif[2.5,4.2]	3.63 (106,891)	3.41 (63,571)	2.73, 4.14
$Log(m_{BO\text{-}ST})$	0.063	unif[-5.0,-3.0]	-3.64	-3.93	-4.72, -3.06
$Log(m_{ST\text{-}BO})$	0.065	unif[-5.0,-3.0]	-3.53	-3.89	-4.65, -3.01
$Log(m_{NT\text{-}ST})$	0.328	unif[-5.0,-3.0]	-3.76	-3.90	-4.66, -3.09
$Log(m_{ST\text{-}NT})$	0.175	unif[-5.0,-3.0]	-3.57	-3.85	-4.58, -3.01
$Log(m_{BO})$	0.248	unif[-4.0,-2.0]	-2.47	-2.87	-3.64, -2.01
$Log(m_{NT})$	0.178	unif[-4.0,-2.0]	-3.17	-3.00	-3.84, -2.15
MALEMIG	0.204	unif[0,1]	0.87	0.70	0.34, 1.00
Log(STR_MUT)	0.832	unif[-5.0,-3.0]	-3.94	-3.95	-4.24, -3.66

[a], BO=Borneo, NT=Sumatra north of Lake Toba, ST=Sumatra south of Lake Toba (a description of the different model parameters can be found in Figure 4.5); [b], coefficient of determination; [c], 90%-highest posterior density interval.

4.4.4. Validation

The validation of the model selection procedure revealed a misclassification rate of 63% over all eight tested demographic models (Supporting Figure S4.5). The high misclassification rate

can be attributed to the broad prior distributions used for the different demographic models (Supporting Table S4.2), which seem to produce strongly overlapping distributions of summary statistics. However, misassignments were mostly confined to the two groups of models (2-population and 10-population models), as the correct group was inferred in 82% of all pseudo-observed data sets.

The selected 10-population model with bottleneck showed a good fit to the observed genetic data, with a p-value of the observed genetic data under the inferred GLM of 0.567. We also tested if our parameter estimation was systematically biased for certain parameters given our choice of summary statistics. The distribution of posterior quantiles within which the pseudo-observed dataset fell did not significantly deviate from the expectation of uniformity for most parameters (Supporting Figure S4.6). Only the posterior distribution for the microsatellite mutation rate appears to be estimated too conservatively, as indicated by a concentration of data points in the center of the histogram.

4.5. Discussion

Our modeling approach using multiple genetic marker systems on a large set of geographically well-defined samples revealed novel insights into the demographic history of orangutans. We showed that gene flow between the two orangutan species ceased ~100 kya, and that this gene flow was predominantly male-mediated. Furthermore, we found evidence for a strong bottleneck on Borneo ~61 kya, which we link to climatic changes during the last glacial period 110-10 kya (Martinson et al. 1987). We could, however, not confirm previous reports of a larger current N_e on Sumatra as compared to Borneo (Steiper 2006; Locke et al. 2011). Our results indicate that such misleading signals are the result of a recent massive decline and deep divergence of orangutan populations on Sumatra, which yields a larger long-term N_e for Sumatran orangutans as compared to Bornean orangutans in oversimplified two-population models.

The selected 10-population model revealed a population split time between Borneo and south of Lake Toba of ~900 kya with subsequent gene flow. This population split estimate between Bornean and Sumatran orangutans is considerably older that the estimates obtained recently using whole genome data (Locke et al. 2011; Mailund et al. 2011), which resulted in split time estimates between 330 and 400 kya (equaling 410–500 kya when assuming a generation time of 25 years as in this study). However, population split times are extremely difficult to estimate in models with migration, since old splits with higher levels of gene flow will produce similar genetic signals as younger splits with less subsequent gene flow (Nielsen & Wakeley 2001; Hey 2006; Becquet & Przeworski 2007). The comparatively young estimate of 334 kya by Mailund et al. (2011) can be attributed to model they applied, which did not allow for migration at all. Such recent population split estimates are in disagreement with findings from studies based on mitochondrial DNA (Xu & Arnason 1996; Zhi et al. 1996; Warren et al. 2001; Zhang et al. 2001; Steiper 2006; Nater et al. 2011), which yielded divergence time estimates of island specific mtDNA lineages of 1–5 Ma. This discrepancy in divergence time estimates can be explained by two factors. First, due to the pronounced philopatric tendencies of female orangutans (Galdikas 1995; Morrogh-Bernard et al. 2011; Arora et al. 2012; Nietlisbach et al. 2012; van Noordwijk et al. 2012), mitochondrial markers

will estimate population split times with little confounding effects of migration. Furthermore, in the absence of female-mediated migration, the coalescence time of island-specific mitochondrial lineages is expected to predate the population split between Borneo and Sumatra, depending on N_e in the ancient population (Nichols 2001).

Due to the combined use of male-inherited Y-chromosomal and female-inherited mtDNA markers, we were for the first time able to confirm that long-term, long-distance migration in orangutans is strongly male-biased, as evidenced by a proportion of males in the migrant pool of ~0.87. Thus, the estimate of a split time between Borneo and Sumatra of around 900 kya in combination with subsequent male-mediated gene flow between the two islands best explains the currently observed genetic patterns in orangutans. While similar genetic patterns of male-biased dispersal in orangutans were found over comparatively short time-spans and geographical scales (Arora et al. 2010; Arora et al. 2012; Nietlisbach et al. 2012), this study is the first to show such an effect over long distances and evolutionarily significant time spans. Our findings indicate that male orangutans distributed genes over large distances, therefore limiting the differentiation potential of local populations. Given the presence of land bridges between Borneo and Sumatra during more than half of the time in the last 250,000 years (Voris 2000) and potentially to a similar extent since the time of the species split, the dispersal ability of male orangutans might in fact provide a good explanation why the two orangutan species have not yet reached reproductive isolation (Muir et al. 1998), despite the long time since population divergence.

The existence of gene flow between Bornean and Sumatran orangutans after the population split is in agreement with previous findings (Muir et al. 2000; Verschoor et al. 2004; Becquet & Przeworski 2007; Locke et al. 2011). Contrary to studies that hinted at the presence of impassable dispersal barriers on the exposed Sunda shelf, either due to large river systems (Harrison et al. 2006) or a putative savannah corridor (Gathorne-Hardy et al. 2002; Bird et al. 2005), it seems that habitat conditions during glacial periods did at least sporadically allow male orangutans to cross the exposed Sunda shelf. However, given the strict and long lasting separation of mtDNA lineages (Nater et al. 2011) on both islands, it appears that the exposed shelf was not covered with rainforest able sustain orangutan populations over prolonged periods. In fact, recent findings show large parts of the Sunda shelf between Borneo and Sumatra being covered with nutrient-poor sandy soils (Bird et al. 2005; Cannon et al. 2009; Slik et al. 2011). Forests on such soil types are characterized by low growth and productivity (Paoli et al. 2010). These constraints might explain why orangutan populations on both islands could not expand onto the exposed shelf to an extent where population admixture and thus exchange of mtDNA lineages was possible.

Since the best fitting model (IM10-BN-DECSU, Figure 4.5) included cessation of gene flow between Borneo and Sumatra at a certain point in time, we were able to infer that migration between the two islands ended ~107 kya. Interestingly, Locke et al. (2011) mentioned that their simple isolation with migration model could not completely explain the relatively common occurrence of shared low-frequency mutations between the islands, and suggested that time-dependent migration models might provide a better fit to the data. In their selected demographic model, the historical migration rates may have been underestimated due to the possibility of migration up to the present time. However, this assumption is unfeasible given

the complete absence of any land bridges between Borneo and Sumatra since at least 10,000 years (Voris 2000). Realistic models of orangutan demographic history should therefore take the dynamics of Pleistocene sea level changes explicitly into account, which will allow us to better understand the dynamics of inter-island gene flow and speciation in orangutans and other Sundaland species.

Our point estimate for the time of cessation of gene flow between Borneo and Sumatra at 107 kya is considerably older than the final flooding of land bridges between the two islands at the end of the last glacial period ~10 kya (Voris 2000). The apparent lack of inter-island gene flow during the last glacial period 110–10 kya, despite the prolonged presence of connecting land bridges, points to additional migration barriers on the exposed shelf during this timeframe. Since the climate during glacial periods was considerably cooler and drier, low sea levels were also associated with a reduction in rainforest coverage (Flenley 1998; Morley 2000; Gathorne-Hardy et al. 2002; Bird et al. 2005), preventing forest dwelling species from dispersing over the exposed Sunda shelf (Gorog et al. 2004). Thus, the windows of opportunity for male orangutans to migrate between Borneo and Sumatra might have been restricted to the initial and final phases of glacial periods, when sea levels where low enough to expose land bridges and climatic conditions allowed for a quick expansion of suitable rainforest habitat onto the exposed shelf. Based on our results, we cannot completely exclude the occurrence of gene flow in the time frame between the last glacial maximum (LGM) 19–26 kya (Clark et al. 2009) and the disappearance of land connections between Borneo and Sumatra ~10 kya, as such a short window of migration might not have left a strong enough signal in our data.

A recent genomic study has shown that Sumatran orangutans exhibit genome-wide higher genetic diversity and corresponding long-term N_e as compared to the Bornean sister species (Locke et al. 2011), despite their nearly ten times lower current census size (Wich et al. 2008). Locke and colleagues explained this somewhat surprising result with a demographic model that included an exponential growth since the population split for Sumatran orangutans and an exponential decline for the Bornean species. Our results indicate that such a simple 2-population model with exponential growth cannot sufficiently explain the current patterns of genetic diversity, as more complex models fit the observed data significantly better. Using a model that incorporated population substructure, we demonstrated that Sumatran orangutans are genetically deeply structured. Furthermore, all Sumatran populations suffered a drastic recent decline and currently exhibit very low effective population sizes. In contrast, Bornean orangutans went through a severe bottleneck during the last glacial period from which they have subsequently expanded. Thus, simplified models that treat Sumatran and Bornean orangutans each as a single panmictic population and do not allow for sudden population size changes will result in N_e estimates that do not properly reflect current census sizes within local subpopulations.

By applying a demographic model that took population structure within both orangutan species into account, we found two signals for a bottleneck on Borneo. First, we demonstrated that the currently observed pattern of strong population differentiation on Borneo (Warren et al. 2001; Arora et al. 2010) has only been established recently, since our selected model indicates that Bornean orangutans were organized at least temporarily as a single panmictic

population before 22 kya. Second, at ~61 kya, the ancient population on Borneo experienced a sudden drop in N_e from ~6,800 to ~650 individuals, which then recovered again ~22 kya to the current total N_e of ~5,300 for all Bornean orangutans. Such a change in both N_e as well as population structure could be explained by a common Bornean refugium during the coldest phase of the last glacial period, when the drier and more seasonal climate might have caused a drastic reduction of rainforest coverage on Borneo (Morley 2000; Gathorne-Hardy et al. 2002; Bird et al. 2005). The population recovery and structuring then coincides with the end of the LGM 19–26 kya. The warmer and wetter climate during this time would have led to an expansion of rainforests and thus suitable orangutan habitat, and might also have led to a strong increase in water flow levels of major river systems on Borneo. Rivers have been shown to be important dispersal barriers for orangutans, and are an important contributor in shaping population structure in orangutans on both Borneo (Goossens et al. 2005; Jalil et al. 2008; Arora et al. 2010) and Sumatra (Nater et al. 2013).

The time estimate for a bottleneck on Borneo is considerably younger than previous findings, which were based on the rapid coalescence of mitochondrial lineages around 150–200 kya (Arora et al. 2010; Nater et al. 2011). Such a bottleneck has therefore been linked with the penultimate glaciation 190-130 kya rather than the last glacial period 110-10 kya. If our conclusion of a common refugial population on Borneo during the last glacial period is correct, then it seems that this most recent bottleneck was not severe and long lasting enough to lead to the coalescence of all mitochondrial lineages currently present in the Bornean populations. Alternatively, female-specific population structure might have persisted during the most recent bottleneck, as the strong philopatric behavior of female orangutans might have prevented complete panmixia in the refugial population. Such population contractions with subsequent expansions probably occurred multiple times on Borneo during Pleistocene glacial and interglacial cycles, explaining why our estimate of the timing of the bottleneck is at odds with those proposed by previous genetic studies (Steiper 2006; Arora et al. 2010; Nater et al. 2011). Supporting this notion, our rather low estimate of N_e in the ancient population on Borneo (~7,000 individuals), despite strong indications of historically much higher than current census numbers of orangutans on Borneo (Rijksen & Meijaard 1999; Meijaard et al. 2010), might be a signal for such recurrent bottlenecks during Pleistocene glaciations.

Linking bottleneck signals to specific environmental processes is difficult due to the large confidence intervals associated with most parameter estimates. For instance, the 90%-highest posterior interval for the estimate of the start of the bottleneck on Borneo (13–287 kya) also overlaps with the Toba supereruption on northern Sumatra ~73 kya (Rose & Chesner 1987; Chesner et al. 1991). It has been hypothesized that this colossal explosive eruption had such a strong impact on the global climate that it caused a severe bottlenecks in humans (Rampino & Ambrose 2000; Ambrose 2003). Still, evidence presented here point toward climatic changes during the last glacial period rather than the Toba supereruption being the main cause for a bottleneck on Borneo. Our results show that the supereruption did not have a strong impact on the Sumatran populations despite their much closer geographic proximity, as signals of population decline on Sumatra were considerably younger than the Toba supereruption. However, a contraction of rainforests during the last glacial period as cause for the bottleneck on Borneo might explain the absence of a similar bottleneck in the Sumatran population

history. During the generally drier glacial periods, large parts of Sumatra experienced considerably more rain fall compared to Borneo (Newsome & Flenley 1988; Whitten et al. 2000; Gathorne-Hardy et al. 2002), because the Barisan mountain range running the length of Sumatra acted as a barrier for the wet monsoon winds, causing high precipitation along its western slopes (Whitten et al. 2000). This mountain ridge effect in combination with the close proximity to the sea during glacial periods with their associated low sea levels might have allowed large areas of rainforest to persist on Sumatra during glacial periods (Gathorne-Hardy et al. 2002). Because of this, Sumatran orangutans were likely not forced into glacial refugia to the same extent as Bornean orangutans.

Interestingly, a similar signal of a glacial refugium with subsequent population structuring, as observed in Bornean orangutans, has been found in western gorillas (*Gorilla gorilla*). By using a demographic modeling approach comparable to our study, Thalmann et al. (2011) found that the two subspecies of western gorillas (*G. g. gorilla* and *G. g. diehli*) diverged only about ~18 kya, thus directly following the LGM 19–26 kya (Clark et al. 2009). Furthermore, the ancient population of western gorillas exhibited a N_e of just ~2,500 individuals as compared to 22,000 and 17,000 individuals in the two subspecies after the population split. Thus, it seems that western gorillas, similar to Bornean orangutans, were constrained to a relatively small refugial population during the last glacial period from which they subsequently expanded when the climate got warmer and wetter after the LGM, leading to the currently observed division into two distinct populations.

While Sumatran orangutans did not seem to go through glacial bottlenecks as the Bornean species, we found evidence for recent and drastic declines in population sizes north and south of Lake Toba. These signals of population decline cannot be attributed to the large-scale human-induced habitat degradation that started in the last century (Rijksen & Meijaard 1999), of which genetic signals may have been found in a previous study in Bornean orangutans (Goossens et al. 2006a). Rather, our results point toward an earlier decline in the Late Pleistocene or Early Holocene. In the Late Pleistocene, orangutans went extinct on the Southeast Asian mainland as well as in many Sundaland regions (Jablonski 1998; Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). Furthermore, the Pleistocene-Holocene boundary is characterized by the disappearance of many large-bodied animals world-wide (Martin & Wright 1967; Martin & Klein 1984; Koch & Barnosky 2006), including large parts of the megafauna in Southeast Asia (Louys et al. 2007). The increased occurrence of megafaunal extinctions during this period has been attributed to climatic changes following the LGM, the impact of human hunting and habitat alterations, or the combination of these two factors (reviewed in Koch & Barnosky 2006).

Both climatic and anthropogenic factors might have played a role in the decline and local extinctions of orangutan populations in the Late Pleistocene. During the LGM, the drier and more seasonal climate caused a shifting of zones of evergreen rainforest toward the equator (Flenley 1998; Jablonski 1998; Morley 2000), likely causing populations in southern China to go extinct. The warmer climate following the LGM was accompanied by rising sea levels, which drastically increased the extent of coastlines in Sundaland (Voris 2000). This enlargement of coastal habitat might have promoted an expansion of early modern humans on Sundaland, leading to increased hunting pressure on large-bodied animals, including

orangutans (Hill et al. 2007; Soares et al. 2008). Such hunting by modern humans might have caused the local extinctions of orangutans on many Sundaland islands, and led to a strong decline in Sumatran populations north and south of Lake Toba. Bornean orangutans did not seem to be as strongly affected by human hunting, probably because the large size and low productivity of Borneo left enough inland areas with relatively low human densities (Delgado & Van Schaik 2000).

Our modeling approach revealed a complex demographic history of orangutans, characterized by deep population splits, inter-island migration, varying degrees of population structure and large population size changes. The two orangutan species, however, experienced completely different demographic scenarios. Sumatran orangutans exhibit a deep and temporally stable population structure, including an old divergence of gene pools north and south of Lake Toba with limited amount of gene flow over the Toba caldera. Therefore, our results confirm previous findings of a strong separating effect of the Toba volcano (Nater et al. 2011; Nater et al. 2013), which showed four major eruptions during the last 1.2 million years (Chesner et al. 1991). The populations on Sumatra recently suffered a strong decline, which, in combination with the strong population structure, explains the high residual genetic diversity present on Sumatra despite the low census size. In contrast, we find that the population structure currently observed within Bornean orangutans is relatively recent and the population went through at least one bottleneck most likely associated with a glacial refugium.

The demographic history of orangutans was heavily influenced by climate changes and anthropogenic pressures, which played an important role in shaping the observed patterns of DNA variation. Thus, our findings have major implications for ongoing studies trying to identify genomic signals of adaptive evolution in great apes in general and orangutans in particular. Using comparable ABC methods, signals of similarly complex demographic histories have been found in humans (Fagundes et al. 2007), chimpanzees (Wegmann & Excoffier 2010) and gorillas (Thalmann et al. 2011). Fagundes and colleagues evaluated the likelihood of different models of human evolution and found strong statistical support for an out-of-Africa replacement model that incorporated strong population bottlenecks during the colonization of Asia and America. Working with chimpanzees (*Pan* spp.), Wegmann and Excoffier inferred that central chimpanzees (*P. t. troglodytes*) are the oldest population within *Pan troglodytes*, since the other two subspecies (*P. t. schweinfurthii* and *P. t. verus*) went through severe bottlenecks when splitting off from the central population. Thalmann and colleagues used a combination of historic and contemporary DNA samples of western gorillas (*Gorilla gorilla*) to show that the Cross River gorilla subspecies (*G. g. diehli*) went through a strong and recent population decline, which the authors directly link to anthropogenic pressure.

These results strongly suggest that special consideration needs to be given to demographic factors when analyzing adaptive evolutionary processes, at least in great apes. Due to their strong dependence on intact forest habitat, most great ape taxa were severely affected by the last glacial period, which was accompanied by drastic changes in forest coverage in the tropics (Flenley 1998; Morley 2000). Accordingly, great ape populations experienced population bottlenecks, founder events, population expansions and population structuring as recent as 15,000 years ago (Clark et al. 2009). Given the long generation time of all great apes

(18–30 years, Wich et al. 2009a), great ape populations will likely not have reached a equilibrium state for most genomic regions. Thus, population expansions and population structuring caused by relatively recent climatic changes might produce erroneous signals of positive selection if demography is not taken into account. Additionally, all great apes experienced and are still experiencing massive recent declines due to anthropogenic pressures, both hunting and habitat degradation (IUCN 2012). Such recent reductions in effective population sizes will lead to genomic signals that might easily be misinterpreted as signals of balancing selection. Given such complex demographic histories, an approach where genomic data is simulated under a refined demographic model as inferred here will allow obtaining a neutral expectation of DNA variation and therefore help to detect deviations due to selection with higher confidence as compared to a stationary equilibrium model.

4.6. Acknowledgements

We thank Pirmin Nietlisbach, Nicole Ponta, Livia Gerber, Corinne Ackermann and Kai Ansmann for providing valuable labwork for this study. We are grateful to Erik Willems, who provided the map of Sundaland. Ian Singleton, Kristin S. Warren, Ernst J. Verschoor, Benoit Goossens, Laurentius N. Ambu, Maria A. van Noordwijk, Helen Morrogh-Bernard, Noko Kuze, Tomoko Kanamori, Joko Pamungkas, and Dyah Perwitasari-Farajallah all provided orangutan samples that were analyzed for this study. A special thank goes to Daniel Wegmann for his technical help with the ABCtoolbox software. This project was financially supported by the Swiss National Science Foundation (grant no. 3100A-116848 to MK and CvS), Forschungskredit of the University of Zurich (grant no. 57020601), Messerli Foundation, A.H.-Schultz Foundation and Claraz Schenkung. Furthermore, we thank the following institutions for supporting our research: Primate Research Center of the Bogor Agricultural University (IPB), Indonesian State Ministry of Research and Technology (RISTEK), Indonesian Institute of Sciences (LIPI), Sabah Wildlife Department, Taman National Gunung Leuser (TNGL), Borneo Orangutan Survival Foundation (BOSF), Leuser International Foundation (LIF), and Badan Pengelola Kawasan Ekosistem Leuser (BPKEL).

Chapter 5: General Discussion

5.1. The Population History of Orangutans

Orangutans have evoke large interest among scientists studying human evolution due to their basal position in the great ape phylogeny, their exclusive occurrence in Southeast Asia, as well as many behavioral and morphological traits that are distinct among great apes, such as the adaptations for arboreality and highly male-biased dispersal patterns (Delgado & Van Schaik 2000). Despite this interest in the evolutionary history of orangutans, knowledge of their population history is still extremely limited. Previous genetic studies have suffered from incomplete sampling coverage, unreliable sample provenance, a restricted number of genetic markers, and mostly qualitative assessments of phylogeographic patterns. The work presented here was aimed at remedying these issues by improving genetic research of orangutans in three main directions: sampling from different wild populations throughout the extant range of both species, use of a large variety of genetic marker systems, as well as the use of modeling frameworks to quantitatively interpret the observed genetic patterns.

My approach allowed us for the first time to describe the genetic variation in both extant orangutan species on mitochondrial, Y-chromosomal and autosomal levels. Based on this multi-locus approach, four main findings emerged. First, we revealed pronounced genetic structure within and between both orangutan species, hinting at strong dispersal barriers. Second, there is a marked contrast in genetic structure between female- and male transmitted markers, indicative of a strong male-bias in dispersal. Third, we confirmed previous findings of a population bottleneck on Borneo (Steiper 2006; Arora et al. 2010), which we link to a refugium during the last glacial period (110–10 kya). In contrast, Sumatran orangutans showed a remarkably stable population history, which only recently underwent a severe decline. Fourth, we found that the only extant orangutan population south of Lake Toba, Batang Toru, shows a deep divergence from the Sumatran orangutans north of Lake Toba. Even though genetic signals of such a distinct southern Sumatran population had been found previously (Muir et al. 2000), they could not be interpreted in a phylogeographic context due to the lack of samples with reliable provenance.

Due to its geographic location south of Lake Toba, the population of Batang Toru is of special interest in investigating the population history of orangutans. Batang Toru was likely part of a larger ancestral meta-population that inhabited central and south Sumatra up to as little as 100 years ago (Rijksen & Meijaard 1999). This ancestral meta-population might have regularly exchanged migrants with Bornean populations during Pleistocene glacial periods, since a land bridge between Sumatra and Borneo was in place around the islands of Bangka and Belitung, near the southeastern coast of Sumatra, at sea levels just 40 meters below present (Rijksen & Meijaard 1999; Voris 2000). Given the deep divergence of populations north and south of Lake Toba (Chapter 3), genetic signals of past gene flow between Borneo and Sumatra are likely better preserved in the Batang Toru population as compared to populations north of Lake Toba. Thus, incorporating genetic samples from the Batang Toru area allowed us to

improve our understanding of the gene flow patterns between Borneo and Sumatra, which is critical to understand the processes of divergence and speciation in the evolutionary history orangutans.

Evidence for such gene flow between Borneo and Sumatra after the population split remains a contentious issue. In most previous studies, mitochondrial data suggested that the populations on Borneo and Sumatra were strictly separated for over a million years (Xu & Arnason 1996; Zhi et al. 1996; Warren et al. 2001; Zhang et al. 2001). Nonetheless, a few studies found genetic signals of recent migration between Bornean and Sumatran orangutans, but these studies were all severely hampered by their exclusive reliance on Sumatran samples from captive individuals. Even though some of these zoo individuals were born in the wild, detailed geographic provenance is missing and alternative explanations of supposed gene flow signals, such as human translocations, could not be ruled out completely. Both Muir et al. (2000) and Kanthaswamy et al. (2006) identified predominantly Bornean haplotypes in orangutans of Sumatran origin, a result they attributed to recent migration from Borneo to Sumatra. Further evidence for recent migration between the two orangutan species was put forward by Verschoor et al. (2004) working with simian foamy viruses. The authors of this study found that some Sumatran orangutans exhibit virus lineages that cluster with lineages found in Bornean orangutans, even though these foamy viruses are expected to have coevolved with their hosts since the initial population split.

More recently, Locke et al. (2011) used complete genome sequences to model the demographic history of orangutans. By testing the fit of nine relatively simple demographic models to the observed DNA polymorphism patterns, the authors favored a model with gene flow up to the present time after the initial population split 400 kya, as well as a continuous exponential population growth on Sumatra and a continuous exponential decline on Borneo after the split. However, a scenario of continuous and ongoing gene flow is unfeasible, given the absence of land bridges since the end of the last glacial period ~10 kya (Voris 2000). Furthermore, the authors did not test more complex demographic models, such as models with the possibility for multiple changes in population size instead of a continuous exponential growth or decline on each island. The relatively good fit of the most complex within a set of simple demographic models does not in any way imply a good absolute fit, which would indicate that a realistic demographic model has been chosen. In fact, Locke et al. (2011) admitted that their selected demographic model was not able to reproduce certain aspects of the observed genetic data, such as the high proportion of low-frequency polymorphisms shared between Bornean and Sumatran orangutans. A more complex demographic model might have resulted in a better overall fit to the observed genetic data. Thus, gross oversimplification might have led to misleading results (Nielsen & Beaumont 2009), as evidenced by their finding of a larger current effective population size in Sumatran than Bornean orangutans, despite the much larger current census sizes of the latter (Wich et al. 2008).

Our results confirm the earlier findings of gene flow between Borneo and Sumatra after the population split (Chapter 4). This is illustrated by the gene trees of four non-coding autosomal regions, where there is paraphyly and even haplotype sharing between the two species using samples with reliable geographic provenance. In agreement with this finding, the most likely

point of interruption of gene flow between Bornean and Sumatran orangutans was at ~107 kya, as inferred by a modeling approach (Chapter 4). Hence, our results do not support the occurrence of gene flow between the two orangutan species during most of the last glacial period 110–10 kya, despite the prolonged presence of land bridges between Borneo and Sumatra (Voris 2000). Rather, our findings indicate that the two extant orangutan species were completely reproductively isolated for the last 100,000 years, but underwent regular genetic exchange during glacial periods before the ultimate one (Chapter 4). The confirmation of relatively recent genetic exchange between Bornean and Sumatran orangutans has implications for the taxonomy of orangutans, which I will discuss in detail below.

In addition to population splits and gene flow patterns, changes in effective population sizes will strongly influence the evolutionary history of a taxon, since effective population size is a major factor in determining the relative importance of selection and genetic drift in populations (Gillespie 2004). Fluctuations in the effective size of a population are mainly linked to changes in the population census size, the rate of genetic exchange with neighboring populations or the population substructure (Nichols et al. 2001), as other factors that determine effective population size in a population, such as social organization, mating system or life history traits, show a high temporal stability within a given species (Frankham 1995). Such changes in population census size, connectivity and substructure can have multiple causes, such as a reduction, fragmentation or expansion of suitable habitat, diseases, predator-prey interactions or geological events. Given the history of the Sunda region, the most likely reasons for sudden changes in effective population sizes are associated with climate and sea level changes during Pleistocene glaciations, the devastating Toba supereruption ~73 kya and, more recently, anthropogenic factors such as hunting and large-scale habitat alteration (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). A recent decline caused by large-scale human activities starting at the beginning of the last century has been well documented (e.g. Rijksen & Meijaard 1999; van Schaik et al. 2001) and signals of this have also been detected with genetic methods (Goossens et al. 2006a). Yet, it is much harder to find evidence for old demographic changes, because such genetic signals are erased by more recent demographic changes (Hudson 1990).

Previous studies found genetic signals for a population bottleneck followed by an expansion on Borneo (Steiper 2006; Arora et al. 2010). However, such signals were not supported by the recent genomic study of Locke et al. (2011), who in contrast identified a continuous population decline in Bornean orangutans since the population split from Sumatrans. However, the simple demographic model applied by Locke and colleagues tried to describe the changes in effective population size on Borneo during the last 400,000 years with a single rate of continuous change, which might have resulted in the mixing of recent genetic signals of a population decline (Goossens et al. 2006a) with older signals of a population expansion. The output of such a simplified model is therefore hard to interpret, and a multitude of demographic events might have led to the finding that a continuous decline on Borneo is better fitting the genetic data than the equally unrealistic options of a continuous expansion or a constant population size. Surprisingly, by using mitochondrial DNA, both Steiper (2006) and Arora et al. (2010) observed signals from a relatively old expansion (70–180 kya) on the background of a strong recent decline. This is remarkable, as the effective population size of mtDNA is around four times lower than that of the autosomal genomic sequences used by

Locke et al. (2011) and signals of older demographic events are expected to be erased quickly in the small current subpopulations. The presence of strong signals of a relatively old Bornean bottleneck in the genealogy of mitochondrial lineages can be explained by their much stronger geographical structuring compared to autosomal and Y-chromosomal markers, caused by pronounced philopatry of female orangutans (Morrogh-Bernard et al. 2011; Arora et al. 2012; Nietlisbach et al. 2012; van Noordwijk et al. 2012). Due to the absence of female-mediated gene flow among differentiated local subpopulations, at least one distinct mitochondrial lineage per subpopulation is retained, and genetic signals of demographic changes are conserved over long time spans, independent of subsequent declines on the subpopulation level (Wakeley & Aliacar 2001).

Recent declines in population sizes and distribution area caused by anthropogenic pressures are well documented in both orangutan species (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000; van Schaik et al. 2001; Goossens et al. 2005; Goossens et al. 2006a; Meijaard et al. 2010). In agreement with these earlier findings, we discovered strong signals for population declines in Sumatran orangutans, but not so for the Bornean species (Chapter 4). We estimated that the most likely time for the decline in Sumatran orangutans falls in the Late Pleistocene for population north of Lake Toba (~15 kya), or Early Holocene (~9 kya) for populations south of Lake Toba. Thus, Sumatran orangutan populations already underwent a decline before large-scale human habitat degradation started at the beginning of the last century (Rijksen & Meijaard 1999). This result is consistent with the complete disappearance of orangutans in many areas of Sundaland in the Late Pleistocene, which has been attributed to prehistoric hunting by early humans (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000; van Schaik et al. 2001). In contrast, the finding of a stable population size on Borneo during the last 20,000 years indicates that Bornean populations, compared to orangutans on Sumatra, were reacting in a different way to the climatic and anthropogenic changes at the end of the last glacial period.

The absence of genetic signals of more recent population declines in our demographic modeling does not imply that orangutans are not negatively affected by the large-scale and ongoing deforestation on the island during the last 100 years (Rijksen & Meijaard 1999). Such recent genetic signals are probably not yet obvious due the extremely long generation time of orangutans (Wich et al. 2009a; Peter et al. 2010). Despite such a long generation time, by Goossens et al. (2006a) found strong genetic signals of dramatic declines during the last 200 years in orangutan populations in Sabah, Borneo. The discrepancy between our study and the findings of Goossens and colleagues can be explained by their higher spatial sampling density, but might also be an artifact of recent immigration from differentiated neighboring populations into the study area (Nielsen & Beaumont 2009; Chikhi et al. 2010; Peter et al. 2010).

5.2. Influences of Environmental Processes on the Population History

Combining the results of genetic analyses with knowledge from the fields of behavioral ecology, geography, geology and climate history allows us to disentangle the factors that have shaped the current patterns of genetic variation in orangutans. Such factors include both species-specific traits such as social organization, ecological requirements and dispersal

behavior, as well as extrinsic factors, such as environmental processes, geographical features and anthropogenic influences. In the following paragraphs, I will focus the discussion on the extrinsic factors, as my work capitalizes on the use of samples from the entire distribution of both orangutan species, which allows us to investigate such processes on a sufficiently large spatial and temporal scale. Intrinsic determinants of genetic variation in orangutans have already been analyzed and discussed in detail in population-level studies (Goossens et al. 2006b; Arora et al. 2012; Nietlisbach et al. 2012), which better allowed them to disentangle the fine-scale patterns of genetic variation, given their higher spatial sampling density.

Extrinsic factors shaping the currently observed distribution of genetic diversity in orangutans include fluctuating sea levels, changes in rainforest coverage and connectivity, volcanic activities, human hunting and habitat destruction, and dispersal barriers such as rivers and mountain ranges. Such processes have been previously investigated in orangutans (e.g. Muir et al. 2000; Warren et al. 2001; Goossens et al. 2005; Goossens et al. 2006a; Steiper 2006; Jalil et al. 2008; Arora et al. 2010). However, these analyses were in many cases restricted to just one species or one local population of orangutans, and thus might have missed important processes that affect the two orangutan species differently. Such differential influences of extrinsic processes might explain many of the striking genetic and phenotypic differences that can be observed between Bornean and Sumatran orangutans (reviewed in Wich et al. 2009b). Figure 5.1 summarizes the current knowledge about the demographic changes in the evolutionary history of orangutans and the environmental processes that might have caused these changes.

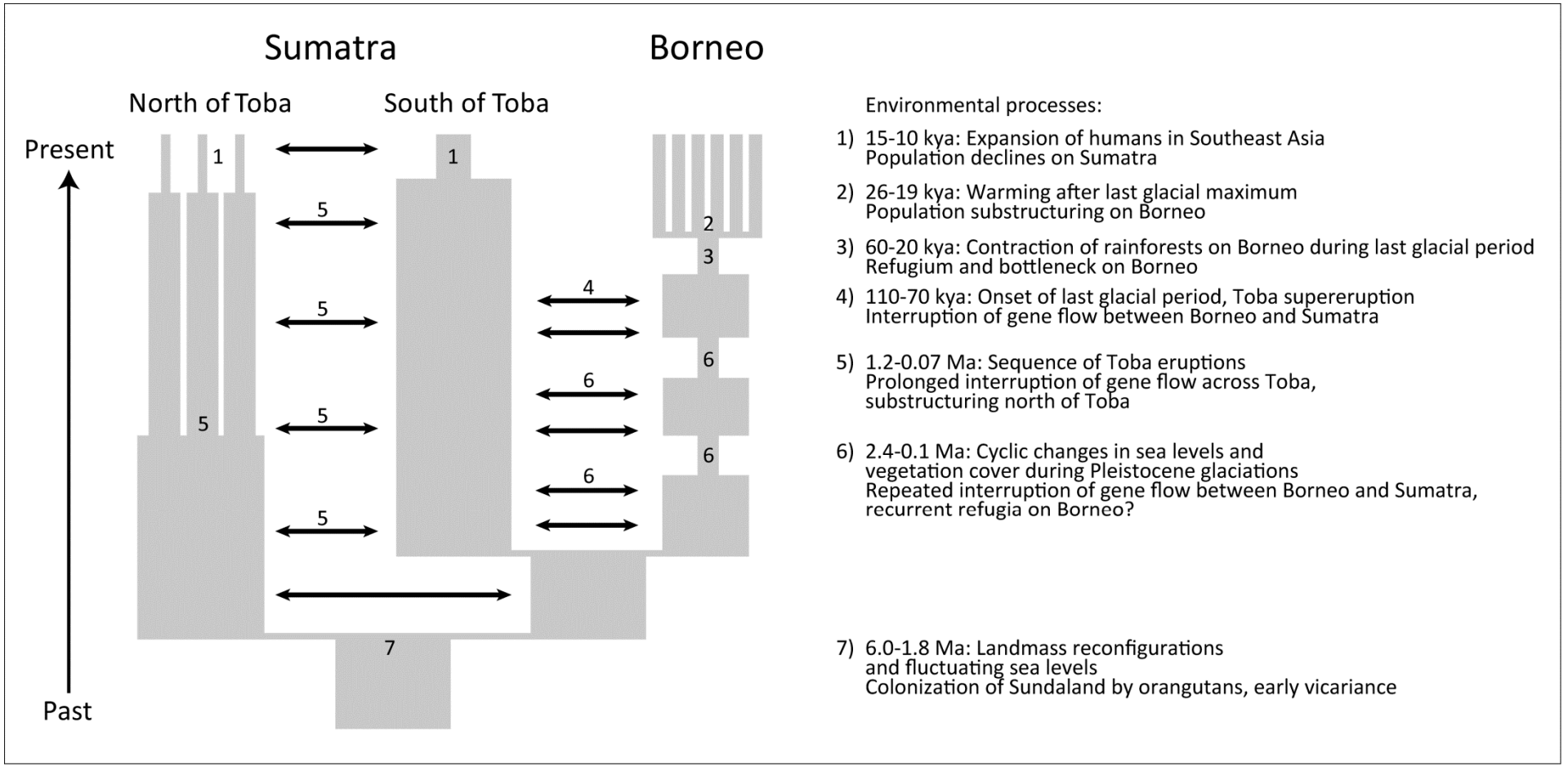


Figure 5.1: Schematic representation of the demographic history of orangutans. Marked demographic changes are linked by numbers to potential underlying environmental processes. Only extant orangutan populations are shown, as demographic information about extinct populations on the mainland and Java is almost completely absent. The time axis is not to scale.

5.2.1. Factors Affecting Migration Opportunities

Given the unique insular habitat of orangutans, changes in sea levels during the Pleistocene are likely a major contributor in shaping genetic patterns in this taxon. The sea in between the Southeast Asian Sunda islands is relatively shallow, and the continental shelf was exposed during recurrent glacial periods in the Pleistocene (Martinson et al. 1987; Voris 2000). Thus, the emergence of land bridges during glacial periods might have allowed land animals to disperse over the exposed Sunda shelf. During interglacials, sea levels rose to a level similar or even above present (Linsley 1996; Voris 2000), which then repeatedly interrupted gene flow for prolonged periods and may have led to distinct biogeographic patterns in many Sundaland taxa, including primates (Brandon-Jones 1996; Groves 2001; Meijaard 2004; Meijaard & Groves 2004).

The climate during glacial periods was generally colder, drier and more seasonal, which caused strong fluctuations in the type and extent of vegetation coverage in the tropics (Flenley 1998; Morley 2000). Migration opportunities of forest-dwelling species in Sundaland therefore largely depended on the habitat conditions on the exposed shelf. Some paleoecological reconstructions of the last glacial period point at a broad savannah corridor on the Sunda shelf (Heaney 1991; Gathorne-Hardy et al. 2002; Bird et al. 2005; Wurster et al. 2010), while other studies hinted at extensive lowland forest coverage throughout Sundaland (Sun et al. 2000; Cannon et al. 2009; Wang et al. 2009). An open savannah type vegetation on the exposed shelf would have prevented dispersal of forest-dwelling species among the Sundaland islands, but the extent of such a savannah corridor might have varied extensively, both within and among glacial periods. In fact, the conflicting evidence found by various paleoecological studies might have been caused by the short time frame around the glacial maxima during which such a savannah corridor might have been present, while the exposed shelf might have been forested during warmer and wetter phases of glacial periods.

A recent genetic study working with three rainforest murine species in Sundaland supported the notion of a strong dispersal barrier for forest-dwelling species on the exposed Sunda shelf, as island-specific mtDNA lineages showed deep divergence and signals of Pleistocene gene flow among islands were completely missing (Gorog et al. 2004). Our results also speak strongly against extensive lowland rainforest coverage on the Sunda shelf during glacial periods. Mitochondrial lineages on Borneo and Sumatra show a deep divergence around 2 Ma, and signals for female-mediated gene flow during subsequent glacial periods are completely absent (Chapter 2). This result indicates that habitat conditions did not allow orangutan populations to expand on the exposed shelf to an extent where populations from Borneo and Sumatra would have met, as such population admixture should have led to the exchange of island-specific mtDNA lineages. Thus, given the deep divergence of mitochondrial lineages in many rainforest-dwelling species on Sundaland (Tosi et al. 2003; Gorog et al. 2004; Ziegler et al. 2007; Arora et al. 2010; Thinh et al. 2010; Nater et al. 2011), habitat conditions on the exposed shelf must have been rather hostile for such species.

In contrast to the deep divergence of mitochondrial lineages, Y-chromosomal markers indicate regular male-mediated gene flow between Borneo and Sumatra as recent as the penultimate glaciation 190–130 kya (Chapter 2). The occurrence of male-mediated inter-island gene flow in such a rainforest dependent taxon as orangutans strongly implies that

forested dispersal corridors on the exposed Sunda shelf existed and that a savannah corridor on the shelf was either spatially or temporally not continuous during glacial periods.

Land connections between Borneo and Sumatra were in place at sea levels just 40 meters below present , while sea levels dropped by up to 120 meters during the last glacial maximum (LGM) (Voris 2000). This left large time windows at the onset and end of glacial periods, when climatic conditions might have been favorable enough to allow exposed land bridges to become temporally forested. Given the sandy and nutrient poor character of topsoil on the shelf between Borneo and Sumatra (Bird et al. 2005; Cannon et al. 2009; Slik et al. 2011), such forest would show extremely low productivity and would probably not be able to sustain permanent orangutan populations. It may, however, allow dispersing male orangutans to cross the exposed shelf and carry genes from Borneo to Sumatra or vice-versa. Indeed, in similarly unproductive forests on Borneo and Sumatra, orangutan males have been observed far away from any permanent population (Rijksen & Meijaard 1999).

Another reason why orangutan males were able to cross a putative savannah corridor might be given by the major rivers systems that dissected the exposed Sunda shelf during glacial periods (Voris 2000; Harrison et al. 2006). These rivers might have provided important dispersal corridors by supporting gallery forests suitable for orangutans in an otherwise rather dry landscape. Thus, a combination of drier and more seasonal climatic conditions during glacial periods, nutrient poor topsoil on the exposed shelf, gallery forests along river banks, and strong philopatric tendencies of female orangutans can explain why orangutan males were able to carry genes between populations on Borneo and Sumatra up to relatively recently, while mitochondrial markers retained a divergence of island-specific lineages of over 2 million years (Chapter 2). Therefore, the striking differences in phylogeographic patterns between mitochondrial and Y-chromosomal markers found in orangutans are the results of both environmental and socio-behavioral processes.

The demographic reconstructions in Chapter 4 indicate that gene flow between the two orangutan species finally ceased at the beginning of the last glacial period ~110 kya. Thus, migration of orangutans between Borneo and Sumatra was likely not possible during most of the last glacial period 110–10 kya, even though this period was characterized by prolonged phases of low sea levels, which completely exposed the Sunda shelf between the two islands (Voris 2000). Given the availability of land connections between Borneo and Sumatra during the last glacial period, additional factors must have determined the dispersal possibilities for forest-dwelling species on the Sunda shelf, and such factors must have varied among different glacial periods. Interestingly, genetic evidence for migration events between major Sundaland islands during the last glacial period in other forest-dwelling species is also strikingly absent, and multiple genetic studies dealing with phylogenetic patterns of such taxa hinted at genetic isolation of island populations considerably older than the start of the last glacial period (Gorog et al. 2004; Quek et al. 2007; Wilting et al. 2007).

Unfortunately, most reconstructions of the paleolandscape of the exposed Sunda shelf focused on the last glacial period, which makes it difficult to identify the factors determining migration possibilities during glacial sea level low stands. A well-documented environmental factor might have contributed to the absence of gene flow across the Sunda shelf during the last glacial period. Around 73 kya, the Toba supereruption on northern Sumatra expelled

colossal amounts of ash and sulphur gases into the upper atmosphere (Rose & Chesner 1987; Chesner et al. 1991; Rampino & Self 1992). This event likely caused an accelerated cooling and transition to glacial conditions on a global scale (Rampino & Self 1992; Williams et al. 2009). On a regional scale, the eruption might have had a severe impact on the rainforest coverage in South Asia, as evidenced by pollen records showing a prolonged period of deforestation in India (Williams et al. 2009). Thus, while the results of our studies do not support a significant impact of the Toba supereruption on the orangutan populations *per se*, the long-term effects of this event on the vegetation in Southeast Asia might have influenced the connectedness of populations on different Sunda islands.

5.2.2. Factors Influencing Population Structure

Volcanic activities also likely caused the most striking phylogeographic signal identified in orangutans. On Sumatra, populations north and south of Lake Toba show an extraordinary deep divergence for mitochondrial markers (Chapter 2). Interestingly, for mtDNA markers, the only extant population south of Lake Toba is phylogenetically more closely related to all Bornean orangutans than to their conspecifics north of Lake Toba. Lake Toba is the water filled caldera that was formed in the process of four major eruptions of the Toba volcano during the last 1.2 million years (Chesner et al. 1991), with the largest being the Toba supereruption ~73 kya. The eruptions of the Toba volcano were strong enough to devastate large areas of rainforest around the caldera (Rampino & Self 1992; Williams et al. 2009), which would have temporarily imposed an insuperable barrier for forest-dwelling species. In between these eruptions, the rainforest seems to have recovered to an extent where dispersal of male orangutans was possible, since there is no similar pattern of deep divergence over Lake Toba for Y-chromosomal or autosomal markers. Our modeling approach confirmed the presence of such a deep population split within Sumatran orangutans and pointed at low levels of male-mediated gene flow between populations north and south of Lake Toba.

The phylogeographic patterns in Sumatran orangutans help to understand some striking species boundaries that occur around Lake Toba without any other obvious biogeographic causes. For example, speciation patterns in other forest-dwelling primate species seem to be heavily influenced by the Toba eruptions. The distributions of Thomas' leaf monkeys (*Presbytis thomasi*) (Aimi & Bakar 1996) and whitehanded gibbons (*Hylobates lar*) in Sumatra are confined to areas north of Lake Toba, and the mountain agile gibbon (*Hylobates agilis*) (Whittaker et al. 2007; Thinh et al. 2010) occurs in Sumatra only south of Lake Toba.

On a more regional scale, rivers have been shown to have a strong impact on the partitioning of genetic diversity in great apes, including orangutans (Eriksson et al. 2004; Goossens et al. 2005; Anthony et al. 2007; Jalil et al. 2008; Arora et al. 2010; Gonder et al. 2011). Major rivers are known to delineate both species and subspecies distributions in great apes (Groves 2001; Gonder et al. 2006; Anthony et al. 2007), indicating a long-lasting separation effect on local gene pools. This is seemingly also the case with orangutans on Borneo, where subpopulations exhibit strong differentiation for both mitochondrial and autosomal microsatellite markers across large rivers (Goossens et al. 2005; Jalil et al. 2008; Arora et al. 2010). However, by looking at longer time scales, we showed that this differentiation has been established only since the LGM (~22 kya, Chapter 4). The drier and more seasonal climate during Pleistocene glacial periods (Flenley 1998; Morley 2000) might have

temporally reduced large rivers to a size that would allow frequent crossing by dispersing individuals, therefore allowing for a homogenization of local gene pools on Borneo. The strong genetic differentiation currently observed on Borneo was then driven by small subpopulation sizes when Bornean orangutans were expanding from a bottleneck during the last glacial period (Chapter 4).

5.2.3. Factor Affecting Changes in Effective Population Size

Demographic modeling revealed large fluctuations in population sizes for both Bornean and Sumatran orangutans, although the patterns were strikingly different between the two orangutan species (Chapter 4). Bornean orangutans have experienced a severe bottleneck 60-20 kya, which is consistent with a refugium on Borneo during the last glacial period. The climatic changes during Pleistocene glaciations strongly influenced the extent of rainforest coverage on Borneo. During recurrent glacial periods, the climate was colder and drier and the rainforest on the island was likely restricted to small refugia (Flenley 1998; Morley 2000; Gathorne-Hardy et al. 2002; Bird et al. 2005). Such a large-scale and prolonged reduction in suitable rainforest habitat is likely the cause for the genetic bottleneck signals documented in Chapter 4 of this thesis, as well as in previous genetic studies (Steiper 2006; Arora et al. 2010). The cyclic nature of Pleistocene climate changes made it highly likely that such bottlenecks reoccurred multiple times during the evolutionary history of Bornean orangutans, although genetic signals of older events might have been erased by the most recent bottleneck. The power to detect genetic signals of such recurrent Pleistocene refugia depends on the length and severity of the last bottleneck as well as the effective size of the genetic marker system used to investigate such patterns (e.g. Ramakrishnan et al. 2005).

Genetic signatures that might be indicative of Pleistocene glacial refugia have been found in other great apes, such as eastern gorillas (Jensen-Seaman & Kidd 2001), western gorillas (Thalmann et al. 2011), bonobos (Fischer et al. 2006; Wegmann & Excoffier 2010), central (Fischer et al. 2004; Fischer et al. 2006), western (Wegmann & Excoffier 2010) and eastern chimpanzees (Goldberg & Ruvolo 1997; Wegmann & Excoffier 2010). The widespread occurrence of such signals in the evolutionary history of great apes points to low ecological tolerance to habitat changes associated with Pleistocene glacial cycles. However, it should be pointed out that most genetic studies working with demographic history in great apes did not employ a comparable modeling approach with consideration of fine-scale population structure as performed in this thesis. Therefore, alternative explanations for genetic signals of population expansions, such as population subdivision or unrepresentative population sampling, cannot be excluded (Wakeley & Aliacar 2001; Ptak & Przeworski 2002).

In contrast to orangutans on Borneo, Sumatran orangutans do not seem to have suffered from glacial bottlenecks, as our results indicate a stable population history for prolonged periods. This result is not surprising, as northern Sumatra has seen more constant rainfall patterns through glacial cycles as compared to Borneo, due to the proximity to the deep sea and the precipitation of wet monsoon winds along the western slope of the Barisan mountain range (Whitten et al. 2000). Sumatran orangutans were therefore expected to be less affected by changes in vegetation cover prevailing on large parts of Sundaland during glacial cycles (Flenley 1998; Morley 2000; Bird et al. 2005). There is, however, strong evidence for a recent decline in Sumatran populations, both north and south of Lake Toba, starting at the

Pleistocene-Holocene boundary 9–15 kya (Chapter 4). This period is well known for dramatic population declines and extinctions in many large-bodied animals world-wide (Martin & Wright 1967; Martin & Klein 1984; Koch & Barnosky 2006), including megafauna in Southeast Asia (Louys et al. 2007). The disappearance of many species during this time frame has been either attributed to climatic changes at the end of the last glacial period or to the impact of human hunting and habitat alterations (reviewed in Koch & Barnosky 2006).

A climatic explanation for a decline in orangutan populations on northern Sumatra at the end of the last glacial period ~15 kya is consistent with a recent study inferring that northern Sumatra, in stark contrast to most other regions on Sundaland, experienced more rainfall during the LGM as compared to today (Kim et al. 2008). Thus, the climatic changes following the LGM 19–26 kya in combination with rising sea levels might have caused a reduction in rainforest habitat on the northwestern part of Sundaland, pushing Sumatran orangutans north of Lake Toba into a refugial state.

In contrast to populations north of Lake Toba, a climatic explanation for Early Holocene population declines is less likely for populations on southern Sumatra, where climatic conditions resembled the situation on Borneo (Whitten et al. 2000; Kim et al. 2008). In these regions of Sundaland, the end of the last glacial period was most likely accompanied by an expansion of rainforest coverage (Flenley 1998; Morley 2000; Gathorne-Hardy et al. 2002; Bird et al. 2005), and such an expansion of suitable orangutan habitat should have led to a population growth on both southern Sumatra and Borneo. Therefore, human hunting provides a viable alternative of why orangutans on southern Sumatra suffered a decline starting in the Early Holocene, eventually resulting in the extinction of nearly all populations south of Lake Toba (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000). The human overhunting hypothesis is supported by findings of large amounts of fossil remains of orangutans in caves in the Padang area in western Sumatra, which were inhabited by early human colonizers (Rijksen 1978). These early modern humans first arrived in Southeast Asia around 50 kya, but lived in relatively low densities and mainly concentrated in coastal areas (Barker et al. 2007). A large demographic and spatial expansion of humans at the end of the last glacial period ~12 kya might have drastically increased the hunting pressure on orangutans and other Sundaland animals (Hill et al. 2007; Soares et al. 2008), leading to strong genetic signals of population declines in the Late Pleistocene and Early Holocene.

In contrast to Sumatran orangutans, we did not find a signal of a Late Pleistocene population decline on Borneo. The lack of genetic signal might be caused by the recurrent occurrence of strong bottlenecks in the Bornean population history, which caused a relatively small long-term effective population size on Borneo (Chapter 4). On the background of such a low effective population size, recent declines would not form as strong a contrast as on Sumatra, where glacial bottlenecks appear to be missing. However, if such declines are linked to human hunting, the impact on Bornean orangutans might have been limited due to two reasons. First, the rapidly expanding rainforest habitat on Borneo after the LGM ~20 kya (Flenley 1998; Morley 2000; Gathorne-Hardy et al. 2002; Bird et al. 2005) might have compensated for any negative impact by allowing expansion of orangutan populations into newly forested areas. Second, early human settlements were likely concentrated along the coastlines, which would have left waste inland areas relatively free of human impact (Soares et al. 2008). Thus,

Bornean orangutans experienced a demographic and spatial expansion following the glacial maximum, and sustained large population sizes until large-scale human forest exploitation started at the beginning of the last century (Rijksen & Meijaard 1999).

The analyses of phylogeographic patterns in orangutans conducted in the framework of this PhD thesis shed light on the influences of dramatic environmental changes on the evolutionary history of the flora and fauna on Sundaland during the last few million years. These processes led to the isolation and divergence of local gene pools, and caused drastic changes in population sizes. Orangutans are strongly adapted to an arboreal lifestyle and therefore depend strongly on an intact rainforest habitat (Rijksen & Meijaard 1999; Delgado & Van Schaik 2000; Wich et al. 2009b). The evolutionary history of orangutans was therefore particularly strongly influenced by environmental factors that affected rainforest coverage in Southeast Asia. The high sensitivity to vegetation changes in combination with an extraordinary slow life history resulted in strong genetic signals of past environmental processes. Thus, insights from the demographic history of orangutans might prove very valuable in the effort to reconstruct the impact of Pleistocene climate changes on the flora and fauna in Southeast Asia.

5.2.4. Implications for the Taxonomy of Orangutans

The taxonomy of orangutans recently underwent a major revision, when the two subspecies of *Pongo pygmaeus* (Bornean orangutan, *P. p. pygmaeus* and Sumatran orangutan, *P. p. abelii*) were elevated to two different species (*Pongo pygmaeus* and *P. abelii*). This decision was mainly based on the results of early genetic studies using mitochondrial data (Ryder & Chemnick 1993; Xu & Arnason 1996; Zhi et al. 1996). Muir et al. (1998) argued strongly against such an elevation of Bornean and Sumatran orangutans to two distinct species based on mitochondrial markers alone. They highlighted two major problems of genetic studies available at that time. First, behavioral observations had hinted at female philopatry and male dispersal in orangutans (Galdikas 1995), and the female-transmitted mitochondrial DNA might therefore not accurately reflect divergence and gene flow between populations on Borneo and Sumatra. Second, the limited genetic sampling in the aforementioned studies did not represent the entire genetic diversity present in both orangutan taxa. The seemingly high genetic differentiation observed between Bornean and Sumatran orangutans might therefore be an artifact of biased sampling of the extremes of the distribution. In fact, morphological data are less conclusive regarding species-level differences between orangutans on both islands, as intra-island variation is as large or larger than inter-island differences for many morphological traits, in particular regarding tooth and cranial morphology (Groves et al. 1992; Uchida 1998; Groves 2001).

Given the exhaustive sampling conducted in the framework of this project, the analysis of multiple genetic marker systems and the use of novel analysis methods, I was able to address most of the open questions regarding genetic differentiation in orangutans. While we confirmed the strict separation and deep divergence of mtDNA lineages present on both islands (Chapter 2), genetic data from Y-chromosomal and autosomal sequences uncovered a different picture. Y-chromosomal markers showed a much younger divergence of island-specific lineages, indicating recent male-mediated gene flow between Borneo and Sumatra (Chapter 2). Four non-coding autosomal sequences revealed paraphyly and haplotypes sharing

between the two islands (Chapter 4), which is also indicative of recent migration between the two islands. In agreement with this finding, Locke et al. (2011) showed that a demographic model incorporating migration between both islands after the population split fits the patterns of genome-wide variation significantly better than a migration-free model. We further investigated these patterns of inter-island migration in our demographic modeling approach, showing that migration between Borneo and Sumatra was predominantly male-driven and ceased ~100 kya (Chapter 4).

The exchange of genes between the gene pools on the two islands as recent as 100 kya, the fact that both orangutan species produce fertile offspring in captivity (Muir et al. 1998), and the widespread presence of paraphyly/polyphyly in autosomal and mitochondrial gene trees indicate that genome-wide genetic differences might be limited. These findings would speak strongly against the recognition of two distinct orangutan species, as Bornean and Sumatran orangutans do not fulfill the criteria of most common species concepts, including the biological and the phylogenetic species concept (reviewed in de Queiroz 1998). However, gene flow levels between the two islands were historically low and repeatedly interrupted for prolonged time spans during interglacial high sea level stands (Linsley 1996; Voris 2000). Such low levels of gene flow (<4 migrants per generation in each direction) are not sufficient to prevent genetic differentiation of neutral loci (Wright 1931; Hartl & Clark 2007). Furthermore, genomic regions under positive selection will behave differently from neutrally evolving loci, depending on the selective advantages of newly arisen alleles (reviewed in Morjan & Rieseberg 2004). Alleles that are strongly favored over a wide range of habitat conditions will spread rapidly within the whole distribution of orangutans, despite low levels of gene flow among local populations. In contrast, alleles conferring selective advantages only in specific environmental conditions will be fixed rapidly in certain populations and eliminated from the gene pools in others. Thus, genomic regions under selection may either show higher or lower levels of local differentiation as compared to neutrally evolving regions. An assessment on whether the two currently recognized orangutan species follow sufficiently different evolutionary trajectories to be considered separate species is therefore impossible based on neutral genetic markers alone.

The presence of profound environmental differences between Bornean and Sumatran rainforest habitats might have allowed Bornean and Sumatran orangutans two evolve habitat-specific adaptations despite regular past gene flow. Many morphological, behavioral and life history traits show marked differences between Bornean and Sumatran orangutans, but the overall pattern of variation indicates the presence of a continuous gradient for most traits, rather than two distinct units (reviewed in van Schaik et al. 2009b). Hopefully, genome-wide scans for island-specific signals of positive selection in orangutans will finally shed some light on the speciation process in these great apes and reveal the adaptive genetic differences between and within Bornean and Sumatran orangutans.

The results of our demographic analyses point at a recent mixing of local gene pools on Borneo in a common refugium during the last glacial period (Chapter 4). Given the strong female philopatric tendencies (Arora et al. 2012; Nietlisbach et al. 2012; van Noordwijk et al. 2012), such a refugial population was most likely not completely panmictic, as mitochondrial data points to a coalescence date of all Bornean haplotypes considerably older than the

supposed bottleneck during the last glacial period (Chapter 2, Arora et al. 2012). A mixing of autosomal gene pools in refugial populations probably took place multiple times during recurrent glacial cycles in the Pleistocene and raises some doubts on adaptive genetic differences among the three recognized subspecies of Bornean orangutans (Groves 2001). It has been postulated that *Pongo pygmaeus morio*, which inhabits the northeastern parts of the island, shows specific adaptations to the challenges posed by the local habitat conditions (van Schaik et al. 2009b). Northeastern Borneo is heavily affected by the El Niño Southern Oscillation (ENSO) phenomenon, which cause severe droughts and forest fires (MacKinnon et al. 1996). Furthermore, it has been shown that ENSO is strongly linked to the temporal occurrence of mast fruiting on this part of Sundaland (Ashton et al. 1988; Wich & Van Schaik 2000). Therefore, fruit availability is temporally extremely unstable and orangutans have to cope with prolonged periods of fruit scarcity. As compared to other Bornean and Sumatran orangutans, *Pongo pygmaeus morio* shows a reduced brain size (Groves 2001; Taylor & van Schaik 2007), a more robust jaw morphology (Taylor 2009), and a shorter interbirth interval (Wich et al. 2009a), which are thought to be specific adaptations to cope with unpredictable and prolonged periods of extremely low energy input (van Schaik et al. 2009b).

A complete homogenization of autosomal gene pools during the last glacial period would imply that local selective adaptations have evolved in relatively short time on Borneo. Such rapid selective differentiation of local gene pools is feasible if selection mostly acted on standing variation in the Bornean gene pool, which would suggest that the bottleneck was not strong and long-lasting enough to eliminate most of the genetic diversity present in the populations on Borneo. This is supported by my finding that gene trees of four autosomal regions show a coalescence date for Bornean lineages much older than the estimated age of the glacial refugium (Chapter 4). Alternatively, phenotypic plasticity might explain some of the morphological and physiological differences that can be observed within orangutans in general and Bornean orangutans in particular.

The phylogeographic and demographic analyses performed in the framework of this project have demonstrated that Sumatran orangutans experienced a different demographic history as compared to their Bornean sister species. We found that Sumatran orangutans are deeply structured into three autosomal genetic clusters, which are delimited by the Toba caldera and the Alas River (Chapter 3). In contrast to Bornean orangutans, we found no signals of bottlenecks or common refugia during glacial periods or in the aftermath of the Toba supereruption ~73 kya, indicating that Sumatran orangutans had a relatively stable population history during the Pleistocene (Chapter 4). However, signals of recent demographic declines in Sumatran orangutans during the Late Pleistocene and Early Holocene are evident, which might have accelerated genetic differentiation of local gene pools due to habitat fragmentation and increased drift effects.

Given the long-lasting separation of Sumatran subpopulations as compared to Bornean subspecies, a revision of the current taxonomic classification might be advisable. Based solely on neutral genetic differentiation, two distinct taxonomic units should be recognized on Sumatra, with the boundary defined by Lake Toba. Even though the Alas River is further subdividing the gene pool north of Lake Toba, genetic signals for recent male-mediated gene flow around the Alas River are evident (Chapter 3), indicating that this river might have been

a much stronger dispersal barrier in the past, when volcanic material from the Toba eruptions dammed the Alas River (van Schaik & Mirmanto 1985). In contrast, demographic modeling indicates that the populations north and south of Lake Toba historically exchanged only a small number of migrants (Chapter 4), which has led to a strong differentiation on the autosomal level (Chapter 3). Furthermore, the Toba caldera marks the deepest split in the mitochondrial phylogeny of all orangutans, with mtDNA lineages north and south of Lake Toba exhibiting a divergence time of ~3.5 million years (Chapter 2). In contrast, the three currently recognized subspecies of Bornean orangutans show an mtDNA divergence of less than 200 kya and even paraphyly/polyphyly for *Pongo pygmaeus pygmaeus* and *P.p. morio*. The four currently recognized subspecies of the common chimpanzee, *Pan troglodytes*, all show a mtDNA divergence of less than 1 Ma, and the subspecies *P. t. troglodytes* forms a paraphyletic group (Gonder et al. 2006). Thus, the neutral genetic differentiation between Sumatran orangutans north and south of Lake Toba is well above what can be observed between subspecies in other great apes.

Unfortunately, little is known so far about differences in physiological and morphological traits within these two distinct genetic units on Sumatra, which could base the definition of taxonomic units within Sumatran orangutans on a broader ecological basis. Assessment of population differentiation based solely on neutrally evolving genetic markers might strongly misrepresent adaptive genetic differences among populations (Morjan & Rieseberg 2004). On the other hand, the survival outlook for many Sumatran orangutan populations is extremely grim (van Schaik et al. 2001; Marshall et al. 2009). Especially the population south of Lake Toba faces a high risk of extinction in the near future due the low estimated census size of just 550 individuals (Wich et al. 2008) and the fact that large parts of the forest in the Batang Toru region currently possess no protected status (Wich et al. 2011). The grim survival outlook for the genetically distinct population south of Lake Toba is a major concern for the long-term conservation of genetic diversity in Sumatran orangutans. Giving the population south of Lake Toba a separate taxonomic status from the populations in the north might raise public awareness for the need to protect orangutan habitat outside the well-known Gunung Leuser National Park and might allow saving these distinct orangutans from extinction.

5.3. Challenges of Reconstructing Demographic History

The patterns of DNA variation within and between species that we observe today are the result of demographic, selective and random processes during the evolutionary history of these species (Hahn et al. 2002; Wall et al. 2002; Haddrill et al. 2005; Stajich & Hahn 2005). Therefore, understanding the evolutionary forces that have shaped these patterns requires profound knowledge about the demographic history of a taxon. Unfortunately, the reconstruction of demographic history based on genetic data is extremely difficult due to multiple reasons. First, the underlying processes shaping genetic variation have a large random component, which incorporates a lot of noise into any genetic data set. Consequently, large amounts of genetic data from multiple independent loci are required to filter out sufficiently strong signals of demographic processes. Second, demographic reconstructions need to be based on completely neutrally evolving loci in order to extract the pure demographic component of DNA variation from genetic data. This is a difficult criterion to

meet, as positive selection will affect large genomic region by hitch-hiking (Smith & Haigh 1974). Third, demographic histories of natural populations can be extremely complex. In order to investigate such processes, simplified demographic models need to be applied. Yet, oversimplified models might produce misleading results if important aspects of demography are ignored (Nielsen & Beaumont 2009). Fourth, genetic signals of past demographic events are irrecoverably overwritten by more recent events, and therefore reconstruction of demographic events gets harder the older the time of interest (Hudson 1990). Finally, multiple demographic scenarios can result in very similar genetic signals. For example, immigration into a population can be easily misinterpreted as a recent population decline, as both processes would lead to old coalescent events of lineages within a population (Nielsen & Beaumont 2009; Chikhi et al. 2010).

Despite recent methodological advances such as Approximate Bayesian Computation (ABC), the desire to use relatively complex demographic models in combination with large amounts of genetic data is still largely hampered by the availability of analytical tools and computational resources. Thus, a reasonable compromise needs to be found, employing a demographic model that incorporates all significant demographic processes, and using enough genetic data to obtain reasonable parameter estimates. The genetic data set should compromise sequence data from multiple independent autosomal loci, which should be long enough to contain enough information about the genealogical history of lineages, but short enough to make recombination events within the sequence stretches extremely unlikely. Furthermore, these loci need to be carefully selected in order to minimize the possibility of hitch-hiking due to nearby regions under selection. As shown in other studies, a set of 30–50 autosomal regions, each covering 0.5–2 kb, can provide good parameter estimation in relatively complex demographic models (Fagundes et al. 2007; Wegmann & Excoffier 2010; Veeramah et al. 2011). The generation of such large amounts of genetic data in a geographically representative sample set with sufficiently large sample sizes per population to accurately estimate allele frequencies (~20 sampled chromosomes per population, Wakeley 2009) requires high quality DNA samples. Non-invasively collected fecal and hair samples are usually highly degraded, which limits the length of amplifiable DNA sequences to usually <500 bp and requires careful cross validation of results to avoid PCR artifacts (Morin et al. 2001). Thus, demographic reconstructions need to rely to some extent on samples from captive and rehabilitant individuals, for which detailed geographic provenance is often not available or unreliable.

The conclusions from many previous studies dealing with population history and demography of orangutans were severely hampered by the use of samples from captive individuals with unknown sample provenance (e.g. Muir et al. 2000; Zhang et al. 2001; Verschoor et al. 2004; Steiper 2006; Locke et al. 2011). My project aimed at rectifying these issues by acquiring DNA samples directly from wild populations throughout the current distribution of orangutans. This strategy provided us with a large collection of geographically well defined DNA samples, but due to the protected status of orangutans, genetic sampling was limited to non-invasive sample collection methods. As my own sampling efforts showed, encounter rates in areas where orangutans occur at low densities are extremely low, and shed hair from night nests are often the only viable sampling strategy to obtain meaningful sample sizes. My laboratory work on such shed hair samples demonstrated, that such kind of genetic material is

not suitable to generate the kind and extent of genetic data required to reconstruct a complex demographic history. Due to the degraded nature of such non-invasively collected samples, genetic analysis was limited to the amplification of mitochondrial markers and short autosomal and Y-chromosomal microsatellites. However, the great value of this and a related project by Natasha Arora lies in the fact that, albeit limited, genetic analysis of a large set of samples from wild populations allowed us to create a detailed database which links genetic traits with geographic provenance. Based on the combined information of both mitochondrial and Y-chromosomal haplotype as well as the genotype of autosomal microsatellites, we can assign every captive wild-born orangutan with high reliability to its natal geographic region. This achievement will allow the future use of high-quality samples from a large number of wild-born captive and rehabilitant orangutans, making it possible to use new-generation sequence generation techniques on high quality samples without compromising on sample provenance information. We already demonstrated the validity of such an approach in Chapter 4, where we improved our demographic inferences by generating autosomal sequences from blood samples of rehabilitant individuals, which we previously assigned to geographic regions by referencing to genetic data from wild-sampled orangutans.

While autosomal regions offer the possibility to employ a large amount of genetic markers with almost completely independent genealogical histories, they are not very powerful to infer old migration and population splitting patterns. In models with migration, genetic signals of population split time and migration are always confounded and extremely difficult to disentangle without strong prior knowledge of one of the two factors (e.g. Nielsen & Wakeley 2001; Hey 2006; Becquet & Przeworski 2007). In species with highly sex-biased dispersal like orangutans, sex-linked marker systems offer an elegant way around this problem. The marker systems transmitted by the philopatric sex allows investigating population splitting events without confounding effects of migration and provide therefore information about processes that are extremely hard resolve with autosomal data only. In contrast, markers transmitted through the dispersing sex will show much stronger signals of migration events. Additionally, sex-linked markers in mammals have, under idealized conditions, a four times lower effective population size as compared to autosomal loci. For markers transmitted by the philopatric sex, this low effective population size results in rapid lineage sorting between populations. For markers transmitted by the dispersing sex, the low effective population size lets coalescence times become quickly independent from the population splitting time (Slatkin 1991; Wakeley & Aliacar 2001), thus allowing to accurately measure migration rates without confounding influence of splitting time.

5.4. Outlook

Disentangling the evolutionary forces that have shaped patterns of DNA variation in extant species is difficult, because neutral demographic processes can produce similar genetic signals as adaptive evolution. If there is good knowledge of the demographic history of a taxon, we can use this information to formulate expectations of genetic patterns under strict neutrality, against which observed patterns of DNA variation can be compared. This can be relatively easily achieved by simulating genetic data under an inferred demographic model thousands of times. By performing a large number of simulations, the stochasticity of the

coalescent process is accounted for, as the resulting simulated data will reflect the whole range of variation expected under completely neutral evolution. Thus, thresholds for summary statistics can be defined and genomic data can be scanned for signals of deviations from neutrality (Nielsen 2005). Genomic regions that fall outside the range of data simulated under neutrality can thus be identified as under selection with high confidence.

The results of the demographic analyses performed during this PhD project will help to understand the patterns of DNA variation in orangutans. In the framework of the PhD project by Maja Greminger, large amounts of genomic sequence data have been generated. Given our knowledge of the demographic history of orangutans, this project will succeed in identifying genomic regions that show signals of adaptations in different orangutan populations. Due to the evolutionary proximity to humans, this project will then heavily capitalize on the intensive research on gene function in the human genome. Applying this knowledge to orangutans will allow detecting potentially interesting genes within the genomic regions that deviate from neutrality and link them to phenotypic traits that are favorable under specific habitat conditions. Thus, this approach might shed light on the genetic basis of systematic geographic differences in morphology, physiology and behavior that have been described in orangutans (Wich et al. 2009b). Understanding how selection has produced these phenotypic differences among populations within and between two closely related species will increase our understanding of evolution in general and might reveal important processes during human evolution.

6. References

Aimi M, Bakar A (1996) Distribution and deployment of *Presbytis melalophos* group in Sumatera, Indonesia. *Primates* **37**, 399-409.

Alter SE, Rynes E, Palumbi SR (2007) DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 15162-15167.

Ambrose SH (2003) Did the super-eruption of Toba cause a human population bottleneck? Reply to Gathorne-Hardy and Harcourt-Smith. *Journal of Human Evolution* **45**, 231-237.

Andolfatto P, Przeworski M (2000) A genome-wide departure from the standard neutral model in natural populations of Drosophila. *Genetics* **156**, 257-268.

Anthony NM, Johnson-Bawe M, Jeffery K, et al. (2007) The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20432-20436.

Arora N, Nater A, van Schaik CP, et al. (2010) Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (*Pongo pygmaeus*). *Proceedings of the National Academy of Sciences* **107**, 21376-21381.

Arora N, van Noordwijk MA, Ackermann C, et al. (2012) Parentage-based pedigree reconstruction reveals female matrilineal clusters and male-biased dispersal in nongregarious Asian great apes, the Bornean orang-utans (*Pongo pygmaeus*). *Molecular Ecology* **21**, 3352-3362.

Ashton PS, Givnish TJ, Appanah S (1988) Staggered flowering in the Dipterocarpaceae - new insights into floral induction and the evolution of mast fruiting in the aseasonal tropics. *American Naturalist* **132**, 44-66.

Avise JC (2000) *Phylogeography : the history and formation of species* Harvard University Press, Cambridge, Mass. ; London.

Ayres JM, Clutton-Brock TH (1992) River boundaries and species range size in Amazonian primates. *American Naturalist* **140**, 531-537.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Barker G, Barton H, Bird M, et al. (2007) The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *Journal of Human Evolution* **52**, 243-261.

Beaumont MA, Nielsen R, Robert C, et al. (2010) In defence of model-based inference in phylogeography REPLY. *Molecular Ecology* **19**, 436-446.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035.

Becquet C, Patterson N, Stone AC, Przeworski M, Reich D (2007) Genetic structure of chimpanzee populations. *Plos Genetics* **3**.

Becquet C, Przeworski M (2007) A new approach to estimate parameters of speciation models with application to apes. *Genome Research* **17**, 1505-1519.

Bengtsson BO (1978) Avoiding inbreeding - at what cost? *Journal of Theoretical Biology* **73**, 439-444.

Bergl RA, Vigilant L (2007) Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (*Gorilla gorilla diehli*). *Molecular Ecology* **16**, 501-516.

Bird MI, Taylor D, Hunt C (2005) Environments of insular Southeast Asia during the last glacial period: a savanna corridor in Sundaland? *Quaternary Science Reviews* **24**, 2228-2242.

Boulesteix AL, Strimmer K (2007) Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics* **8**, 32-44.

Brandon-Jones D (1996) The Asian colobinae (Mammalia: Cercopithecidae) as indicators of Quaternary climatic change. *Biological Journal of the Linnean Society* **59**, 327-350.

Brunet M, Guy F, Pilbeam D, et al. (2002) A new hominid from the Upper Miocene of Chad, central Africa. *Nature* **418**, 145-151.

Cannon CH, Morley RJ, Bush ABG (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 11188-11193.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195-205.

Chesner CA, Rose WI, Deino A, Drake R, Westgate JA (1991) Eruptive history of earths largest Quaternary caldera (Toba, Indonesia) clarified. *Geology* **19**, 200-203.

Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA (2010) The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* **186**, 983-U347.

Clark PU, Dyke AS, Shakun JD, et al. (2009) The last glacial maximum. *Science* **325**, 710-714.

Cook SR, Gelman A, Rubin DB (2006) Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics* **15**, 675-692.

Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *Bmc Bioinformatics* **11**.

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation. In: *Endless Forms: Species and Speciation* (eds. Howard DJ, Berlocher SH), pp. pp. 57-75. Oxford University Press, New York, Oxford.

Delgado RA, Van Schaik CP (2000) The behavioral ecology and conservation of the orangutan (*Pongo pygmaeus*): a tale of two islands. *Evolutionary Anthropology* **9**, 201-218.

Dobson FS (1982) Competition for mates and predominant juvenile male dispersal in mammals. *Animal Behaviour* **30**, 1183-1192.

Douadi MI, Gatti S, Levrero F, et al. (2007) Sex-biased dispersal in western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology* **16**, 2247-2259.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**, 699-710.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, -.

Eriksson J, Hohmann G, Boesch C, Vigilant L (2004) Rivers influence the population genetic structure of bonobos (*Pan paniscus*). *Molecular Ecology* **13**, 3425-3435.

Eriksson J, Siedel H, Lukas D, et al. (2006) Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (*Pan paniscus*). *Molecular Ecology* **15**, 939-949.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Excoffier L, Foll M (2011) Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332-1334.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285-298.
Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.
Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.
Fagundes NJR, Ray N, Beaumont M, et al. (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 17614-17619.
Fischer A, Pollack J, Thalmann O, Nickel B, Paabo S (2006) Demographic history and genetic differentiation in apes. *Current Biology* **16**, 1133-1138.
Fischer A, Wiebe V, Paabo S, Przeworski M (2004) Evidence for a complex demographic history of chimpanzees. *Molecular Biology and Evolution* **21**, 799-808.
Flenley JR (1998) Tropical forests under the climates of the last 30,000 years. *Climatic Change* **39**, 177-197.
Frankham R (1995) Effective population-size adult-population size ratios in wildlife - a review. *Genetical Research* **66**, 95-107.
Franklin IR (1980) Evolutionary change in small populations. In: *Conservation biology : an evolutionary-ecological perspective* (eds. Soulé ME, Wilcox BA), pp. 135-149. Sinaver, Sunderland, Mass.
Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.
Galdikas BMF (1995) Social and reproductive behavior of wild adolescent female orangutans. In: *The Neglected Ape* (eds. Nadler RD, Galdikas BFM, Sheeran LK, Rosen N), pp. 163-182. Plenum Press, New York.
Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305-318.
Gascon C, Malcolm JR, Patton JL, et al. (2000) Riverine barriers and the geographic distribution of Amazonian species. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13672-13677.
Gathorne-Hardy FJ, Syaukani, Davies RG, Eggleton P, Jones DT (2002) Quaternary rainforest refugia in south-east Asia: using termites (Isoptera) as indicators. *Biological Journal of the Linnean Society* **75**, 453-466.
Gaveau DLA, Wich S, Epting J, et al. (2009) The future of forests and orangutans (*Pongo abelii*) in Sumatra: predicting impacts of oil palm plantations, road construction, and mechanisms for reducing carbon emissions from deforestation. *Environmental Research Letters* **4**.
Gerloff U, Schlotterer C, Rassmann K, et al. (1995) Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). *Molecular Ecology* **4**, 515-518.
Gillespie JH (2004) *Population genetics: a concise guide*, Second Edition edn. The Johns Hopkins University Press, Baltimore.
Goldberg TL, Ruvolo M (1997) The geographic apportionment of mitochondrial genetic diversity in East African chimpanzees, *Pan troglodytes schweinfurthii*. *Molecular Biology and Evolution* **14**, 976-984.
Goldstein DB, Linares AR, Cavallisforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 6723-6727.

Gonder MK, Disotell TR, Oates JF (2006) New genetic evidence on the evolution of chimpanzee populations and implications for taxonomy. *International Journal of Primatology* **27**, 1103-1127.
Gonder MK, Locatelli S, Ghobrial L, et al. (2011) Evidence from Cameroon reveals differences in the genetic structure and histories of chimpanzee populations. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4766-4771.
Goodall J (1983) Population dynamics during a 15 year period in one community of free-living chimpanzees in the Gombe-National-Park, Tanzania. *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology* **61**, 1-60.
Goossens B, Chikhi L, Ancrenaz M, et al. (2006a) Genetic signature of anthropogenic population collapse in orang-utans. *Plos Biology* **4**, 285-291.
Goossens B, Chikhi L, Jalil MF, et al. (2005) Patterns of genetic diversity and migration in increasingly fragmented and declining orang-utan (*Pongo pygmaeus*) populations from Sabah, Malaysia. *Molecular Ecology* **14**, 441-456.
Goossens B, Setchell JM, James SS, et al. (2006b) Philopatry and reproductive success in Bornean orang-utans (Pongo pygmaeus). *Molecular Ecology* **15**, 2577-2588.
Gorog AJ, Sinaga MH, Engstrom MD (2004) Vicariance or dispersal? Historical biogeography of three Sunda shelf murine rodents (*Maxomys surifer*, *Leopoldamys sabanus* and *Maxomys whiteheadi*). *Biological Journal of the Linnean Society* **81**, 91-109.
Graves JAM (2006) Sex chromosome specialization and degeneration in mammals. *Cell* **124**, 901-914.
Groves CP (2001) *Primate taxonomy* Smithsonian Institution Press, Washington, D.C. ; London.
Groves CP, Westwood C, Shea BT (1992) Unfinished business - Mahalanobis and a clockwork orang. *Journal of Human Evolution* **22**, 327-340.
Guilyardi E, Wittenberg A, Fedorov A, et al. (2009) Understanding El Nino in Ocean-Atmosphere General Circulation Models Progress and Challenges. *Bulletin of the American Meteorological Society* **90**, 325-+.
Guschanski K, Caillaud D, Robbins MM, Vigilant L (2008) Females shape the genetic structure of a gorilla population. *Current Biology* **18**, 1809-1814.
Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005) Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Research* **15**, 790-799.
Hahn MW (2008) Toward a selection theory of molecular evolution. *Evolution* **62**, 255-265.
Hahn MW, Rausher MD, Cunningham CW (2002) Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. *Genetics* **161**, 11-20.
Hall R (1996) Reconstructing Cenozoic SE Asia. In: *Tectonic Evolution of Southeast Asia* (eds. Hall R, Blundell D), pp. 153-184. The Geological Society, London.
Handley LJL, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology* **16**, 1559-1578.
Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution* **16**, 700-706.
Harpending HC, Batzer MA, Gurven M, et al. (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1961-1967.
Harrison RD (2001) Drought and the consequences of El Nino in Borneo: a case study of figs. *Population Ecology* **43**, 63-75.

Harrison T, Krigbaum J, Manser J (2006) Primate biogeography and ecology on the Sunda shelf islands: a paleontological and zooarchaeological perspective. In: *Primate Biogeography*. Springer US.

Hartl DL, Clark AG (2007) *Principles of population genetics*, 4th edn. Sinauer Associates, Sunderland, Mass.

Hasegawa M, Kishino H, Yano TA (1985) Dating of the human ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174.

Heaney LR (1991) A synopsis of climatic and vegetational change in Southeast Asia. *Climatic Change* **19**, 53-61.

Hedrick PW, Kalinowski ST (2000) Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics* **31**, 139-162.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics & Development* **16**, 592-596.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.

Hill C, Soares P, Mormina M, et al. (2007) A mitochondrial stratigraphy for island southeast Asia. *American Journal of Human Genetics* **80**, 29-43.

Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxford Surveys in Evolutionary Biology* (eds. Futuyma DJ, Antonovics JD), pp. 1-44. Oxford University Press, New York.

Husson SJ, Wich SA, Marshall AJ, et al. (2009) Orangutan distribution, density, abundance and impacts of disturbance In: *Orangutans: geographic variation in behavioral ecology and conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.

Inger RF, Voris HK (2001) The biogeographical relations of the frogs and snakes of Sundaland. *Journal of Biogeography* **28**, 863-891.

IUCN (2012) IUCN red list of threatened species. Version 2012.1. http://www.iucnredlist.org.

Jablonski NG (1998) The response of catarrhine primates to Pleistocene environmental fluctuations in East Asia. *Primates* **39**, 29-37.

Jalil MF, Cable J, Inyor JS, et al. (2008) Riverine effects on mitochondrial structure of Bornean orang-utans (*Pongo pygmaeus*) at two spatial scales. *Molecular Ecology* **17**, 2898-2909.

Jeffreys H (1961) *Theory of probability*, 3rd edn. Clarendon Press, Oxford.

Jensen-Seaman MI, Kidd KK (2001) Mitochondrial DNA variation and biogeography of eastern gorillas. *Molecular Ecology* **10**, 2241-2247.

Johnson ML, Gaines MS (1990) Evolution of dispersal - theoretical models and empirical tests using birds and mammals. *Annual Review of Ecology and Systematics* **21**, 449-480.

Kaessmann H, Wiebe V, Weiss G, Pääbo S (2001) Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nature Genetics* **27**, 155-156.

Kahlke HD (1972) A review of the Pleistocene history of the orang-utan (*Pongo* Lacepede 1799). *Asian Perspectives*, 5-13.

Kano T (1992) *The last ape : pygmy chimpanzee behavior and ecology* Stanford University Press, Stanford, Calif.

Kanthaswamy S, Kurushima JD, Smith DG (2006) Inferring *Pongo* conservation units: a perspective based on microsatellite and mitochondrial DNA analyses. *Primates* **47**, 310-321.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.
Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.
Kim SJ, Crowley TJ, Erickson DJ, et al. (2008) High-resolution climate simulation of the last glacial maximum. *Climate Dynamics* **31**, 1-16.
Kimura M, Weiss GH (1964) Stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561-576.
Kingsley S (1982) Causes of non-breeding and the development of secondary sexual characteristics in the male orangutan: a hormonal study. In: *The orang-utan: its biology and conservation* (ed. de Boer LEM), pp. p. 215–229. Dr W. Junk Publishers, The Hague.
Knott CD (1998) Changes in orangutan caloric intake, energy balance, and ketones in response to fluctuating fruit availability. *International Journal of Primatology* **19**, 1061-1079.
Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology* **11**, 2623-2635.
Koch PL, Barnosky AD (2006) Late quaternary extinctions: state of the debate. *Annual Review of Ecology Evolution and Systematics* **37**, 215-250.
Krützen M, Willems EP, van Schaik CP (2011) Culture and geographic variation in orangutan behavior. *Curr Biol* **21**, 1808-1812.
Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **22**, 768-770.
Lanave C, Preparata G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20**, 86-93.
Langergraber KE, Siedel H, Mitani JC, et al. (2007) The genetic signature of sex-biased migration in patrilocal chimpanzees and humans. *Plos One* **2**, e973.
Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics* **184**, 243-252.
Linsley BK (1996) Oxygen-isotope record of sea level and climate variations in the Sulu Sea over the past 150,000 years. *Nature* **380**, 234-237.
Lisiecki LE, Raymo ME (2005) A Pliocene-Pleistocene stack of 57 globally distributed benthic delta O-18 records (vol 20, art no PA1003, 2005). *Paleoceanography* **20**, -.
Locke DP, Hillier LW, Warren WC, et al. (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* **469**, 529-533.
Louys J (2007) Limited effect of the Quaternary's largest super-eruption (Toba) on land mammals from Southeast Asia. *Quaternary Science Reviews* **26**, 3108-3117.
Louys J, Curnoe D, Tong HW (2007) Characteristics of Pleistocene megafauna extinctions in Southeast Asia. *Palaeogeography Palaeoclimatology Palaeoecology* **243**, 152-173.
MacKinnon J (1974) The behaviour and ecology of wild orang-utans (Pongo pygmaeus). *Animal Behaviour* **22**, 3-74.
MacKinnon K, Hatta G, Halim H, Mangalir A (1996) *The ecology of Kalimantan* Periplus Editions Ltd., Hong Kong.
Macpherson JM, Ramachandran S, Diamond L, Feldman MW (2004) Demographic estimates from Y chromosome microsatellite polymorphisms: analysis of a worldwide sample. *Hum Genomics* **1**, 345-354.
Maddison WP (1997) Gene trees in species trees. *Systematic Biology* **46**, 523-536.
Mailund T, Dutheil JY, Hobolth A, Lunter G, Schierup MH (2011) Estimating divergence time and ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent hidden Markov model. *Plos Genetics* **7**.

Marjoram P, Molitor J, Plagnol V, Tavare S (2003) Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15324-15328.
Marjoram P, Tavare S (2006) Modern computational approaches for analysing molecular genetic variation data. *Nature Reviews Genetics* **7**, 759-770.
Marshall AJ, Lacy R, Ancrenaz M, et al. (2009) Orangutan population biology, life history, and conservation. In: *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.
Martin PS, Klein RG (1984) *Quaternary extinctions : a prehistoric revolution* University of Arizona Press, Tucson, Ariz.
Martin PS, Wright HR (1967) *Prehistoric extinctions: the search for a cause* Yale University Press, New Haven, CT.
Martinson DG, Pisias NG, Hays JD, et al. (1987) Age dating and the orbital theory of the ice ages - development of a high-resolution-0 to 300,000-year chronostratigraphy. *Quaternary Research* **27**, 1-29.
Mayr E (1970) *Populations, species, and evolution* Belknap Press of Harvard University Press ; London : Distributed by Oxford University Press, Cambridge (Mass.).
Meijaard E (2004) *Solving mammalian riddles: a reconstruction of the Tertiary and Quaternary distribution of mammals and their palaeoenvironments in island South-East Asia* Thesis (Ph D), Australian National University.
Meijaard E, Groves CP (2004) The biogeographical evolution and phylogeny of the genus *Presbytis*. *Primate Report* **68**, 71-90.
Meijaard E, Welsh A, Ancrenaz M, et al. (2010) Declining orangutan encounter rates from Wallace to the present suggest the species was once more abundant. *Plos One* **5**.
Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**, 31-46.
Mevik BH, Wehrens R (2007) The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software* **18**.
Mitani JC, Watts DP, Muller MN (2002) Recent developments in the study of wild chimpanzee behavior. *Evolutionary Anthropology* **11**, 9-25.
Molengraaff GAF (1921) Modern Deep-Sea Research in the East Indian Archipelago. *Geographical Journal* **57**, 95-121.
Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology* **10**, 1835-1844.
Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**, 1341-1356.
Morley RJ (2000) *Origin and evolution of tropical rain forests* Wiley, Chichester.
Morrogh-Bernard HC, Husson SJ, Knott CD, et al. (2009) Orangutan activity budgets and diet: a comparison between species, populations and habitats. In: *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.
Morrogh-Bernard HC, Morf NV, Chivers DJ, Krutzen M (2011) Dispersal patterns of orang-utans (*Pongo* spp.) in a Bornean peat-swamp forest. *International Journal of Primatology* **32**, 362-376.
Muir CC, Galdikas BMF, Beckenbach AT (1998) Is there sufficient evidence to elevate the orangutan of Borneo and Sumatra to separate species? *Journal of Molecular Evolution* **46**, 378-379.

Muir CC, Galdikas BMF, Beckenbach AT (2000) mtDNA sequence diversity of orangutans from the islands of Borneo and Sumatra. *Journal of Molecular Evolution* **51**, 471-480.
Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.
Nater A, Arora N, Greminger MP, et al. (2013) Marked population structure and recent migration in the critically endangered Sumatran orangutan (*Pongo abelii*). *Journal of Heredity* **104**, 2-13.
Nater A, Nietlisbach P, Arora N, et al. (2011) Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant orangutans (genus: *Pongo*). *Molecular Biology and Evolution* **28**, 2275-2288.
Nei M (1987) *Molecular evolutionary genetics* Columbia University Press, New York, NY.
Nei M, Suzuki Y, Nozawa M (2010) The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics* **11**, 265-289.
Newsome J, Flenley JR (1988) Late Quaternary vegetational history of the central highlands of Sumatra. II. Paleopalynology and vegetational history. *Journal of Biogeography* **15**, 555-578.
Newton MA, Raftery AE, Davison AC, et al. (1994) Approximate Bayesian-inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B-Methodological* **56**, 3-48.
Nichols R (2001) Gene trees and species trees are not the same. *Trends in Ecology & Evolution* **16**, 358-364.
Nichols RA, Bruford MW, Groombridge JJ (2001) Sustaining genetic variation in a small population: evidence from the Mauritius kestrel. *Molecular Ecology* **10**, 593-602.
Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39**, 197-218.
Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology* **18**, 1034-1047.
Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**, 885-896.
Nielsen R, Williamson S, Kim Y, et al. (2005) Genomic scans for selective sweeps using SNP data. *Genome Research* **15**, 1566-1575.
Nietlisbach P, Arora N, Nater A, et al. (2012) Heavily male-biased long-distance dispersal of orang-utans (genus: *Pongo*), as revealed by Y-chromosomal and mitochondrial genetic markers. *Molecular Ecology* **21**, 3173-3186.
Nietlisbach P, Nater A, Greminger MP, Arora N, Krützen M (2010) A multiplex-system to target 16 male-specific and 15 autosomal genetic markers for orang-utans (genus: *Pongo*). *Conservation Genetics Resources* **2**, 153-158.
Nsubuga AM, Robbins MM, Roeder AD, et al. (2004) Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology* **13**, 2089-2094.
Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* **22**, 201-204.
Paoli GD, Wells PL, Meijaard E, et al. (2010) Biodiversity conservation in the REDD. *Carbon Balance Manag* **5**, 7.
Peter BM, Wegmann D, Excoffier L (2010) Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. *Molecular Ecology* **19**, 4648-4660.
Pontzer H, Raichlen DA, Shumaker RW, Ocobock C, Wich SA (2010) Metabolic adaptation for low energy throughput in orangutans. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14048-14052.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Molecular Biology and Evolution* **16**, 1791-1798.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
Ptak SE, Przeworski M (2002) Evidence for population growth in humans is confounded by fine-scale population structure. *Trends in Genetics* **18**, 559-563.
Pusey A, Wolf M (1996) Inbreeding avoidance in animals. *Trends in Ecology & Evolution* **11**, 201-206.
Pusey AE (1987) Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends in Ecology & Evolution* **2**, 295-299.
Quek SP, Davies SJ, Ashton PS, Itino T, Pierce NE (2007) The geography of diversification in mutualistic ants: a gene's-eye view into the Neogene history of Sundaland rain forests. *Molecular Ecology* **16**, 2045-2062.
Raaum RL, Sterner KN, Noviello CM, Stewart CB, Disotell TR (2005) Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *Journal of Human Evolution* **48**, 237-257.
Ramakrishnan U, Hadly EA, Mountain JL (2005) Detecting past population bottlenecks using temporal genetic data. *Molecular Ecology* **14**, 2915-2922.
Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer.
Rampino MR, Ambrose SH (2000) Volcanic winter in the Garden of Eden: the Toba supereruption and the late Pleistocene human population crash. In: *Volcanic Hazards and Disasters in Human Antiquity* (eds. McCoy FW, Heiken G). Geological Society of America, Boulder.
Rampino MR, Self S (1992) Volcanic winter and accelerated glaciation following the Toba supereruption. *Nature* **359**, 50-52.
Rasmusson EM, Wallace JM (1983) Meteorological Aspects of the El-Nino Southern Oscillation. *Science* **222**, 1195-1202.
Redelings BD, Suchard MA (2005) Joint Bayesian estimation of alignment and phylogeny. *Systematic Biology* **54**, 401-418.
Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology* **17**, 230-237.
Reis KR, Garong AM (2001) Late Quaternary terrestrial vertebrates from Palawan Island, Philippines. *Palaeogeography Palaeoclimatology Palaeoecology* **171**, 409-421.
Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.
Richard M, Thorpe RS (2001) Can microsatellites be used to infer phylogenies? Evidence from population affinities of the Western Canary Island lizard (Gallotia galloti). *Molecular Phylogenetics and Evolution* **20**, 351-360.
Rijksen HD (1978) *A field study on Sumatran orangutans (Pongo pygmaeus abelii Lesson 1827): ecology, behavior and conservation* H. Veenman and Zonen, Wageningen, The Netherlands.
Rijksen HD, Meijaard E (1999) *Our vanishing relative: the status of wild orang-utans at the close of the twentieth century* Springer Netherlands.
Robert CP, Cornuet JM, Marin JM, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15112-15117.
Roman J, Palumbi SR (2003) Whales before whaling in the North Atlantic. *Science* **301**, 508-510.

Rose WI, Chesner CA (1987) Dispersal of ash in the great Toba eruption, 75 ka. *Geology* **15**, 913-917.
Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3**, 380-390.
Ross KG (2001) Molecular ecology of social behaviour: analyses of breeding systems and genetic structure. *Molecular Ecology* **10**, 265-284.
Rubinoff D, Holland BS (2005) Between two extremes: Mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology* **54**, 952-961.
Ryder OA, Chemnick LG (1993) Chromosomal and mitochondrial-DNA variation in orangutans. *Journal of Heredity* **84**, 405-409.
Sathiamurthy E, Voris HK (2006) Maps of holocene sea level transgression and submerged lakes on the Sunda shelf. *The Natural History Journal of Chulalongkorn University* **Supplement 2**, 1-44.
Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365-371.
Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**, 1079-1089.
Schwarz G (1978) Estimating dimension of a model. *Annals of Statistics* **6**, 461-464.
Selander RK (1983) Evolutionary consequences of inbreeding. In: *Genetics and conservation : a reference for managing wild animal and plant populations* (eds. Schonewald-Cox CM, Chambers SM, MacBryde B, Thomas WL), pp. xxii,722p. Benjamin/Cummings, Merle Park ; London.
Setia TM, van Schaik CP (2007) The response of adult orang-utans to flanged male long calls: inferences about their function. *Folia Primatologica* **78**, 215-226.
Singleton I, van Schaik CP (2001) Orangutan home range size and its determinants in a Sumatran swamp forest. *International Journal of Primatology* **22**, 877-911.
Singleton I, van Schaik CP (2002) The social organisation of a population of Sumatran orang-utans. *Folia Primatologica* **73**, 1-20.
Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.
Slatkin M (1991) Inbreeding coefficients and coalescence times. *Genetical Research* **58**, 167-175.
Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457-462.
Slik JWF, Aiba SI, Bastian M, et al. (2011) Soils on exposed Sunda Shelf shaped biogeographic patterns in the equatorial forests of Southeast Asia. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12343-12347.
Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research* **23**, 23-35.
Soares P, Ermini L, Thomson N, et al. (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *American Journal of Human Genetics* **84**, 740-759.
Soares P, Trejaut JA, Loo JH, et al. (2008) Climate change and postglacial human dispersals in Southeast Asia. *Molecular Biology and Evolution* **25**, 1209-1218.
Sodhi NS, Koh LP, Brook BW, Ng PKL (2004) Southeast Asian biodiversity: an impending disaster. *Trends in Ecology & Evolution* **19**, 654-660.
Stadler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P (2009) The Impact of Sampling Schemes on the Site Frequency Spectrum in Nonequilibrium Subdivided Populations. *Genetics* **182**, 205-216.

Stajich JE, Hahn MW (2005) Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution* **22**, 63-73.
Steadman DW (1993) Biogeography of Tongan Birds before and after Human Impact. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 818-822.
Steiper ME (2006) Population history, biogeography, and taxonomy of orangutans (Genus : *Pongo*) based on a population genetic meta-analysis of multiple loci. *Journal of Human Evolution* **50**, 509-522.
Steiper ME, Young NM (2006) Primate molecular divergence dates. *Molecular Phylogenetics and Evolution* **41**, 384-394.
Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978-989.
Storz JF (1999) Genetic consequences of mammalian social structure. *Journal of Mammalogy* **80**, 553-569.
Sun XJ, Li X, Luo YL, Chen XD (2000) The vegetation and climate at the last glaciation on the emerged continental shelf of the South China Sea. *Palaeogeography Palaeoclimatology Palaeoecology* **160**, 301-316.
Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**, 199-203.
Taberlet P, Griffin S, Goossens B, et al. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189-3194.
Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology & Evolution* **14**, 323-327.
Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-460.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.
Tata HL, van Noordwijk M, Mulyoutami E, et al. (2010) Human livelihoods, ecosystem services and the habitat of the Sumatran orangutan: rapid assessment in Batang Toru and Tripa. World Agroforestry Centre (ICRAF) Southeast Asia Regional Office, Bogor, Indonesia.
Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**, 6463-6471.
Tautz D, Gerloff U, Hartung B, Fruth B, Hohmann G (1999) Intracommunity relationships, dispersal pattern and paternity success in a wild living community of Bonobos (*Pan paniscus*) determined from DNA analysis of faecal samples. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 1189-1195.
Tavare S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505-518.
Taylor AB (2009) The functional significance of variation in jaw form in orangutans. In: *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.
Taylor AB, van Schaik CP (2007) Variation in brain size and ecology in *Pongo*. *Journal of Human Evolution* **52**, 59-71.
Team RDC (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? *Genome Research* **16**, 702-712.

Thalmann O, Wegmann D, Spitzner M, et al. (2011) Historical sampling reveals dramatic demographic changes in western gorilla populations. *Bmc Evolutionary Biology* **11**.

Thinh VN, Mootnick AR, Geissmann T, et al. (2010) Mitochondrial evidence for multiple radiations in the evolutionary history of small apes. *Bmc Evolutionary Biology* **10**, -.

Tosi AJ, Morales JC, Melnick DJ (2003) Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution* **57**, 1419-1435.

Trivers RL (1972) Parental investment and sexual selection. In: *Sexual selection and the descent of man, 1871-1971* (ed. Campbell BG), pp. x, 378 p. Aldine Pub. Co, Chicago.

Tsaousis AD, Martin DP, Ladoukakis ED, Posada D, Zouros E (2005) Widespread recombination in published animal mtDNA sequences. *Molecular Biology and Evolution* **22**, 925-933.

Uchida A (1998) Variation in tooth morphology of *Pongo pygmaeus*. *Journal of Human Evolution* **34**, 71-79.

Utami Atmoko SS, Mitra Setia T, Goossens B, et al. (2009a) Orangutan mating behaviour and strategies. In: *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.

Utami Atmoko SS, Singleton I, van Noordwijk MA, van Schaik C, Setia TM (2009b) Male-male relationships in orangutans. In: *Orangutans: geographic variation in behavioral ecology and conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.

Utami SS, Goossens B, Bruford MW, de Ruiter JR, van Hooff JARAM (2002) Male bimaturism and reproductive success in Sumatran orang-utans. *Behavioral Ecology* **13**, 643-652.

van Noordwijk MA, Arora N, Willems EP, et al. (2012) Female philopatry and its social benefits among Bornean orangutans. *Behavioral Ecology and Sociobiology* **66**, 823-834.

van Noordwijk MA, van Schaik CP (2005) Development of ecological competence in Sumatran orangutans. *American Journal of Physical Anthropology* **127**, 79-94.

van Schaik CP (2004) *Among orangutans: red apes and the rise of human culture* Belknap Press of Harvard University Press, Cambridge, Mass. ; London.

van Schaik CP, Ancrenaz M, Borgen G, et al. (2003) Orangutan cultures and the evolution of material culture. *Science* **299**, 102-105.

van Schaik CP, Ancrenaz M, Djojoasmoro R, et al. (2009a) Orangutan cultures revisited. In: *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.

van Schaik CP, Knott CD (2001) Geographic variation in tool use on Neesia fruits in orangutans. *American Journal of Physical Anthropology* **114**, 331-342.

van Schaik CP, Marshall AJ, Wich SA (2009b) Geographic variation in orangutan behavior and biology. In: *Orangutans: geographic variation in behavioral ecology and conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.

van Schaik CP, Mirmanto E (1985) Spatial variation in the structure and litterfall of a Sumatran rain forest. *Biotropica* **17**, 196-205.

van Schaik CP, Monk KA, Robertson JMY (2001) Dramatic decline in orang-utan numbers in the Leuser Ecosystem, Northern Sumatra. *Oryx* **35**, 14-25.

van Schaik CP, Priatna A, Priatna D (1995) Population estimates and habitat preferences of orangutans based on line transects of nests. In: *The Neglected Ape* (eds. Nadler RD, Galdikas BFM, Sheeran LK, Rosen N), pp. pp. 129-147. Plenum Press, New York.

Veeramah KR, Wegmann D, Woerner A, et al. (2011) An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Molecular Biology and Evolution*.

Verschoor EJ, Langenhuijzen S, Bontjer I, et al. (2004) The phylogeography of orangutan foamy viruses supports the theory of ancient repopulation of Sumatra. *Journal of Virology* **78**, 12712-12716.

Verstappen HT (1973) *A geomorphological reconnaissance of Sumatera and adjacent islands (Indonesia)* Wolters-Noordhoff B.V., Groningen.

Vignaud P, Duringer P, Mackaye HT, et al. (2002) Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. *Nature* **418**, 152-155.

von Koenigswald GHR (1982a) Distribution and evolution of the orang-utan, *Pongo pygmaeus* (Hoppius). In: *The orang-utan: its biology and conservation* (ed. de Boer LEM). Dr W. Junk Publishers, The Hague.

von Koenigswald GHR (1982b) Distribution and evolution of the orang utan, *Pongo pygmaeus* (Hoppius). In: *The orang-utan: its biology and conservation* (ed. de Boer LEM). Dr W. Junk Publishers, The Hague.

Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* **27**, 1153-1167.

Wagner A (2008) Neutralism and selectionism: a network-based reconciliation. *Nature Reviews Genetics* **9**, 965-974.

Wakeley J (2009) *Coalescent theory : an introduction* Roberts & Co. Publishers, Greenwood Village, Colo.

Wakeley J, Aliacar N (2001) Gene genealogies in a metapopulation. *Genetics* **159**, 893-905.

Wall JD, Andolfatto P, Przeworski M (2002) Testing models of selection and demography in *Drosophila simulans*. *Genetics* **162**, 203-216.

Wang L, Sarnthein M, Erlenkeuser H, et al. (1999) East Asian monsoon climate during the Late Pleistocene: high-resolution sediment records from the south China Sea. *Marine Geology* **156**, 245-284.

Wang XM, Sun XJ, Wang PX, Stattegger K (2009) Vegetation on the Sunda shelf, South China Sea, during the last glacial maximum. *Palaeogeography Palaeoclimatology Palaeoecology* **278**, 88-97.

Warren KS, Verschoor EJ, Langenhuijzen S, et al. (2001) Speciation and intrasubspecific variation of Bornean orangutans, *Pongo pygmaeus pygmaeus*. *Molecular Biology and Evolution* **18**, 472-480.

Wegmann D, Excoffier L (2010) Bayesian inference of the demographic history of chimpanzees. *Molecular Biology and Evolution* **27**, 1425-1435.

Wegmann D, Leuenberger C, Excoffier L (2009a) *ABCtoolbox manual*.

Wegmann D, Leuenberger C, Excoffier L (2009b) Efficient Approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* **182**, 1207-1218.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *Bmc Bioinformatics* **11**.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Whitmore TC, Burnham CP (1984) *Tropical rain forests of the Far East*, 2nd ed. edn. Clarendon.

Whittaker DJ, Morales JC, Melnick DJ (2007) Resolution of the Hylobates phylogeny: congruence of mitochondrial d-loop sequences with molecular, behavioral, and morphological data sets. *Molecular Phylogenetics and Evolution* **45**, 620-628.

Whitten T, Damanik SJ, Anwar J, Hisyam N (2000) *The ecology of Sumatra* Periplus Editions Ltd., Hong Kong.

Wich S, Riswan, Jenson J, Refish J, Nelleman C (2011) Orangutans and the economics of sustainable forest management in Sumatra. In: *UNEP/GRASP/PanEco/YEL/ICRAF/GRID-Arendal*.

Wich SA, Buij R, van Schaik C (2004) Determinants of orangutan density in the dryland forests of the Leuser Ecosystem. *Primates* **45**, 177-182.

Wich SA, de Vries H, Ancrenaz M, et al. (2009a) Orangutan life history variation. In: *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (eds. Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP). Oxford University Press.

Wich SA, Meijaard E, Marshall AJ, et al. (2008) Distribution and conservation status of the orang-utan (*Pongo* spp.) on Borneo and Sumatra: how many remain? *Oryx* **42**, 329-339.

Wich SA, Utami Atmoko SS, Mitra Setia T, van Schaik CP (2009b) *Orangutans: geographic variation in behavioral ecology and conservation* Oxford University Press.

Wich SA, Van Schaik CP (2000) The impact of El Nino on mast fruiting in Sumatra and elsewhere in Malesia. *Journal of Tropical Ecology* **16**, 563-577.

Williams GC (1966) *Adaptation and natural selection: a critique of some current evolutionary thought* Princeton University Press, Princeton, N.J.

Williams MAJ, Ambrose SH, van der Kaars S, et al. (2009) Environmental impact of the 73 ka Toba super-eruption in South Asia. *Palaeogeography Palaeoclimatology Palaeoecology* **284**, 295-314.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* **150**, 499-510.

Wilson IJ, Weale ME, Balding DJ (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. *Journal of the Royal Statistical Society Series a-Statistics in Society* **166**, 155-188.

Wilting A, Buckley-Beason V, Feldhaar H, et al. (2007) Clouded leopard phylogeny revisited: support for species recognition and population division between Borneo and Sumatra. *Frontiers in Zoology* **4**, 15.

Woodruff DS (2003) Neogene marine transgressions, palaeogeography and biogeographic transitions on the Thai-Malay Peninsula. *Journal of Biogeography* **30**, 551-567.

Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**, 0097-0159.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Wright S (1955) Classification of the factors of evolution. *Cold Spring Harbor Symposia on Quantitative Biology* **20**, 16-D24.

Wurster CM, Bird MI, Bull ID, et al. (2010) Forest contraction in north equatorial Southeast Asia during the Last Glacial Period. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15508-15511.

Xu XF, Arnason U (1996) The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. *Journal of Molecular Evolution* **43**, 431-437.

Yang ZH, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**, 212-226.

Yasuda M, Matsumoto J, Osada N, et al. (1999) The mechanism of general flowering in Dipterocarpaceae in the Malay Peninsula. *Journal of Tropical Ecology* **15**, 437-449.

Zhang YW, Ryder OA, Zhang YP (2001) Genetic divergence of orangutan subspecies (Pongo pygmaeus). *Journal of Molecular Evolution* **52**, 516-526.

Zhi L, Karesh WB, Janczewski DN, et al. (1996) Genomic differentiation among natural populations of orang-utan (Pongo pygmaeus). *Current Biology* **6**, 1326-1336.

Ziegler T, Abegg C, Meijaard E, et al. (2007) Molecular phylogeny and evolutionary history of Southeast Asian macaques forming the M-silenus group. *Molecular Phylogenetics and Evolution* **42**, 807-816.

7. Supporting Material

7.1. Chapter 2

Supporting Table S2.1: List of samples used for analyses

Type	Status[a]	Cat[b]	Species	Region[c]	Origin	Sex	16S[d]	CYTB[d]	ND3[d]	Y[d]
Blood	R	D	*P. abelii*	BT	Sumatra	male	4	9	5	12
Blood	R	A	*P. abelii*	BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	male	4	8	5	
Feces	W		*P. abelii*	BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	male	4	8	5	12
Feces	W		*P. abelii*	BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	n/a	4	8	5	
Feces	W		*P. abelii*	BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	female	4	8	5	
Feces	W		*P. abelii*	BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	male	4	8	5	13
Feces	W		*P. abelii*	BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	female	4	8	5	
Blood	R	C	*P. abelii*	BT	Tukka, Tapanuli Tengah, North Sumatra	female	4	8	5	
Blood	R	A	*P. abelii*	BT	Pinang Sori, Tapanuli Tengah, North Sumatra	male	4			6
Hair	W		*P. abelii*	BA	Perolihen, Pakpak Bharat, North Sumatra	male	1	5	1	2
Hair	R	C	*P. abelii*	BA	Salak, Pakpak Bharat, North Sumatra	female	1	5	1	
Hair	R	B	*P. abelii*	BA	Perolihen, Pakpak Bharat, North Sumatra	male	1	5	1	2
Hair	R	C	*P. abelii*	BA	Singkohor, Aceh Singkil, Aceh	female	1	5	1	
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female	1	4	1	
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female	1	5	1	
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female	1	5	1	
Blood	R	B	*P. abelii*	KE	Lak-Lak, Aceh Tenggara, Aceh	male				6
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male				6
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male				10
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male				4
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male				4
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	1	5	1	3
Feces	W		*P. abelii*	KE	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	1	4	1	6
Blood	R	A	*P. abelii*	KE	Ladang Lawe Bengkari, Aceh Tenggara, Aceh	male	1	6	1	4
Hair	R	B	*P. abelii*	KE	Pindie, Gayo Lues, Aceh	female	1	4	1	
Hair	R	A	*P. abelii*	KE	Lawe Pupus, Aceh Tenggara, Aceh	male	1	5	1	4

Type	Status[a]	Cat[b]	Species	Region[c]	Origin	Sex	16S[d]	CYTB[d]	ND3[d]	Y[d]
Blood	R	A	*P. abelii*	LK	Tangkahan, Langkat, North Sumatra	male	3	7	4	5
Blood	R	C	*P. abelii*	LK	Langkat, North Sumatra	male	3	7	4	11
Blood	R	B	*P. abelii*	LK	Pinding, Gayo Lues, Aceh	female	3	7	4	
Feces	W		*P. abelii*	LK	Ketambe, East Side of Alas River, Aceh Tenggara, Aceh	female	3	7	4	
Feces	W		*P. abelii*	LK	Ketambe, East Side of Alas River, Aceh Tenggara, Aceh	female	3	7	4	
Feces	W		*P. abelii*	LK	Sampan Getek, Langkat, North Sumatra	female	3	7	4	
Feces	W		*P. abelii*	LK	Sampan Getek, Langkat, North Sumatra	male	3	7	4	4
Feces	W		*P. abelii*	LK	Sampan Getek, Langkat, North Sumatra	female	3	7	4	
Feces	W		*P. abelii*	LK	Sampan Getek, Langkat, North Sumatra	male	3	7	4	6
Feces	W		*P. abelii*	LK	Sampan Getek, Langkat, North Sumatra	female	3	7	4	
Feces	W		*P. abelii*	LK	Sampan Getek, Langkat, North Sumatra	female	3	7	4	
Feces	W		*P. abelii*	LK	Sampan Getek, Langkat, North Sumatra	female	3	7	4	
Feces	W		*P. abelii*	LK	Sikundur, Langkat, North Sumatra	female	3	7	4	
Hair	R	A	*P. abelii*	LK	Pinding, Gayo Lues, Aceh	male	3	7	4	6
Blood	R	C	*P. abelii*	NA	Gayo Lues, Aceh	male	2	3	3	4
Blood	R	C	*P. abelii*	NA	Northern Aceh	male	2	3	3	5
Blood	R	D	*P. abelii*	NA	Sumatra	male	2	3	3	6
Blood	R	D	*P. abelii*	NA	Sumatra	female	2	3	3	
Blood	R	D	*P. abelii*	NA	Sumatra	female	2	3	3	
Hair	R	C	*P. abelii*	NA	Takengon, Aceh Tengah, Aceh	female	2	3	3	
Hair	R	C	*P. abelii*	NA	Panton Labu, Aceh Utara, Aceh	male	2	3	3	
Feces	W		*P. abelii*	SQ	Trumon, Aceh Selatan, Aceh	male	1	4	1	
Blood	R	B	*P. abelii*	SQ	Aceh Selatan, near Suaq Balimbing, Aceh	male	1	5	1	4
Blood	R	A	*P. abelii*	SQ	Aceh Selatan, near Suaq Balimbing, Aceh	female	1	4	1	
Blood	R	A	*P. abelii*	SQ	Aceh Selatan, near Suaq Balimbing, Aceh	female	1	5	1	
Blood	R	B	*P. abelii*	SQ	Aceh Selatan, near Suaq Balimbing, Aceh	male	1	1	2	2
Blood	R	A	*P. abelii*	SQ	Aceh Selatan, near Suaq Balimbing, Aceh	female	1	5	1	
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	female	1	5	1	
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male	1	5	1	
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	female	1	5	1	
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	female	1	5	1	
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				7
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male	1	5	1	
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				4
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				8
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				9
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				2

Type	Status[a]	Cat[b]	Species	Region[c]	Origin	Sex	16S[d]	CYTB[d]	ND3[d]	Y[d]
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				7
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				2
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				4
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				2
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male				4
Feces	W		*P. abelii*	SQ	Suaq Balimbing, Aceh Selatan, Aceh	male	1	5	1	4
Blood	R	B	*P. abelii*	TR	Aceh Barat Daya, Aceh	male	1	1	2	2
Blood	R	B	*P. abelii*	TR	Aceh Barat, Aceh	male	1	1	2	3
Hair	R	C	*P. abelii*	TR	Meulaboh, Aceh Barat, Aceh	female	1	2	1	
Hair	R	B	*P. abelii*	TR	Trangon, Gayo Lues, Aceh	female	1	4	1	
Hair	R	C	*P. abelii*	TR	Meulaboh, Aceh Barat, Aceh	female	1	1	2	
Blood	R	C	*P. abelii*	TR	Aluebillie, Aceh Nagan Raya, Aceh	male	1	1	2	1
Hair	R	C	*P. abelii*	TR	Trangon, Gayo Lues, Aceh	male	1	5	1	2
Hair	W		*P. pygmaeus*	EK	Muara Wahau, East Kalimantan	male	8	16	11	39
Hair	W		*P. pygmaeus*	EK	Muara Wahau, East Kalimantan	male				40
Hair	R	C	*P. pygmaeus*	EK	East Kalimantan	male	8	17	11	
Hair	W		*P. pygmaeus*	EK	Muara Wahau, East Kalimantan	male	8	16	11	40
Hair	W		*P. pygmaeus*	EK	Muara Wahau, East Kalimantan	male				40
Hair	W		*P. pygmaeus*	EK	Sangatta, East Kalimantan	n/a	8	16	11	
Hair	W		*P. pygmaeus*	EK	Kutai National Park, East Kalimantan	n/a	8	12	11	
Hair	W		*P. pygmaeus*	EK	Sangatta, East Kalimantan	n/a	8	16	11	
Hair	W		*P. pygmaeus*	EK	Kutai National Park, East Kalimantan	n/a	8	12	11	
Blood	R	B	*P. pygmaeus*	EK	Kutai National Park, East Kalimantan	male				21
Blood	R	B	*P. pygmaeus*	EK	Kutai National Park, East Kalimantan	male				39
Blood	R	B	*P. pygmaeus*	EK	Muara Wahau, East Kalimantan	male				40
Blood	R	B	*P. pygmaeus*	EK	Muara Wahau, East Kalimantan	male				39
Blood	R	B	*P. pygmaeus*	NK	Lahud Datu, Sabah	female	5	12	12	
Blood	R	B	*P. pygmaeus*	NK	Lahud Datu, Sabah	female	5	12	12	
Blood	R	B	*P. pygmaeus*	NK	Lahud Datu, Sabah	female	5	12	12	
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	male				47
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	male				48
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	male				49
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	female	5	12	12	
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	male	5	12	12	50
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	male	5	12	12	49
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	male				49
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	male				48

Type	Status[a]	Cat[b]	Species	Region[c]	Origin	Sex	16S[d]	CYTB[d]	ND3[d]	Y[d]
Feces	W		*P. pygmaeus*	NK	Lower Kinabatangan, North Side, Sabah	male				49
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male				24
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	female	5	12	9	
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male	5	12	6	25
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male	5	12	9	
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male				26
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male				26
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male				26
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male				25
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male				27
Feces	W		*P. pygmaeus*	SA	West Sabangau, Central Kalimantan	male				28
Hair	R	C	*P. pygmaeus*	SR	Sarawak	male	5	11	7	21
Hair	R	C	*P. pygmaeus*	SR	Sarawak	male	5	12	7	22
Hair	R	C	*P. pygmaeus*	SR	Sarawak	female	5	12	7	
Hair	R	C	*P. pygmaeus*	SR	Sarawak	female	6	13	11	
Hair	R	C	*P. pygmaeus*	SR	Sarawak	male				23
Hair	R	C	*P. pygmaeus*	SR	Sarawak	n/a	5	14	7	
Hair	R	C	*P. pygmaeus*	SR	Sarawak	n/a	5	12	7	
Hair	W		*P. pygmaeus*	SR	Meliau, Danau Sentarum, Northwest Kalimantan	n/a	7	12	8	
Hair	W		*P. pygmaeus*	SR	Meliau, Danau Sentarum, Northwest Kalimantan	n/a	5	12	7	
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	male	5	12	12	41
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	female	5	12	12	
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	female	5	12	12	
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	female	5	12	12	
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	male	5	12	12	41
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	male	5	12	12	42
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	male	5	12	12	41
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	male	5	12	12	41
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	female	5	12	12	
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	female	5	12	12	
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	male	5	12	12	43
Feces	W		*P. pygmaeus*	SK	Danum Valley, Sabah	male				44
Feces	W		*P. pygmaeus*	SK	Lower Kinabatangan, South Side, Sabah	male				41
Feces	W		*P. pygmaeus*	SK	Lower Kinabatangan, South Side, Sabah	male	5	12	12	45
Feces	W		*P. pygmaeus*	SK	Lower Kinabatangan, South Side, Sabah	male				41
Feces	W		*P. pygmaeus*	SK	Lower Kinabatangan, South Side, Sabah	male				43
Feces	W		*P. pygmaeus*	SK	Lower Kinabatangan, South Side, Sabah	male				45

Type	Status[a]	Cat[b]	Species	Region[c]	Origin	Sex	16S[d]	CYTB[d]	ND3[d]	Y[d]
Feces	W		*P. pygmaeus*	SK	Lower Kinabatangan, South Side, Sabah	male	5	12	12	46
Feces	W		*P. pygmaeus*	SK	Lower Kinabatangan, South Side, Sabah	male				43
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				28
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				29
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				30
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				29
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				31
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				28
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				24
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	female	5	15	6	
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				32
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male				33
Feces	W		*P. pygmaeus*	SL	Sungai Lading, Central Kalimantan	male	5	15	6	24
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male	5	15	6	35
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	female	5	12	10	
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				36
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				37
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				35
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				24
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	female	5	12	10	
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				37
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				38
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				18
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				35
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				37
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				36
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male				35
Feces	W		*P. pygmaeus*	TU	Tuanan, Central Kalimantan	male	5	12	10	34
Hair	R	B	*P. pygmaeus*	WK	Pontianak, West Kalimantan	male	5	10	6	14
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	female	5	10	6	
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	male	5	10	6	15
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	male				16
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	female	5	10	6	
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	male	5	10	6	16
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	female	5	10	6	
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	male				17
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	female	5	10	6	

Type	Status[a]	Cat[b]	Species	Region[c]	Origin	Sex	16S[d]	CYTB[d]	ND3[d]	Y[d]
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	female	5	10	6	
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	male	5	10	6	18
Feces	W		*P. pygmaeus*	WK	Gunung Palung, West Kalimantan	male				19
Blood	R	B	*P. pygmaeus*	WK	Pontianak, West Kalimantan	male				14
Blood	R	B	*P. pygmaeus*	WK	Pontianak, West Kalimantan	male				20
Blood	R	B	*P. pygmaeus*	WK	Pontianak, West Kalimantan	male				14
Blood	R	B	*P. pygmaeus*	WK	Pontianak, West Kalimantan	male				18

[a], sample status, W=sampled in the wild, R=rehabilitant individual; [b], provenance reliability of samples from rehabilitant individuals, A=area of capture known with high reliability, B=area of capture known but potentially unreliable, C=only area of confiscation known, D=area of confiscation unknown; [c], sampling region used for analyses (see section 3.4.); [d], haplotypes of mtDNA and Y-chromosomal loci.

Supporting Table S2.2: Nucleotide substitution models selected for different data partitions

Partition	AIC[a]	BIC[b]
all mtDNA	TrN+I	HKY+G
all RNA	TrN+I	TrN+I
tRNA	HKY	HKY
rRNA	TrN+G	TrN+G
all coding	TrN+I	TrN+I
ND3	TrN+I	HKY+I
CYTB	HKY+I	HKY+I

[a], nucleotide substitution model selected according to the Akaike information criterion; [b], model selected according to the Bayesian information criterion. Only substitution models supported by the BEAST software are shown.

Supporting Table S2.3: Illustration of the 13 tested partitioning schemes (rows)

	Sequence sets							
	non-coding regions		ND3 codon position			CYTB codon position		
Partitioning scheme	tRNA	rRNA	1st	2nd	3th	1st	2nd	3th
1	A	A	A	A	A	A	A	A
2	A	A	B	B	B	B	B	B
3A	A	A	B	B	C	B	B	C
3B	A	A	B	B	B	C	C	C
3C	A	B	C	C	C	C	C	C
4A	A	A	B	C	D	B	C	D
4B	A	B	C	C	D	C	C	D
4C	A	B	C	C	C	D	D	D
5A	A	A	B	B	C	D	D	E
5B	A	B	C	D	E	C	D	E
6	A	B	C	C	D	E	E	F
7	A	A	B	C	D	E	F	G
8	A	B	C	D	E	F	G	H

The 1355 sites of the mtDNA dataset have been divided into 8 sets according to their functional characteristics (columns). Sequence sets with the same letter code have been united into a common partition using the same model parameters.

Supporting Table S2.4: Log_{10} Bayes factors for all tested partitioning schemes (see Table S2.2)

Scheme	p[a]	ln P(D|M)[b]	S.E.[c]	1	2	3A	3B	3C	4A	4B	4C	5A	5B	6	7	8
1	233	-3796.49	0.16	-	-13.51	-77.47	-13.04	-12.77	-90.71	-75.66	-13.71	-76.81	-89.83	-75.41	-88.88	-88.10
2	242	-3765.38	0.18	13.51	-	-63.95	0.48	0.74	-77.19	-62.14	-0.20	-63.30	-76.31	-61.90	-75.36	-74.59
3A	249	-3618.12	0.19	77.47	63.95	-	64.43	64.70	-13.24	1.81	63.75	0.66	-12.36	2.06	-11.41	-10.63
3B	248	-3766.48	0.19	13.04	-0.48	-64.43	-	0.26	-77.67	-62.62	-0.68	-63.78	-76.79	-62.37	-75.84	-75.06
3C	248	-3767.08	0.20	12.77	-0.74	-64.70	-0.26	-	-77.93	-62.88	-0.94	-64.04	-77.05	-62.64	-76.10	-75.33
4A	**256**	**-3587.63**	**0.21**	**90.71**	**77.19**	**13.24**	**77.67**	**77.93**	**-**	**15.05**	**76.99**	**13.90**	**0.88**	**15.30**	**1.83**	**2.61**
4B	255	-3622.29	0.22	75.66	62.14	-1.81	62.62	62.88	-15.05	-	61.94	-1.16	-14.17	0.25	-13.22	-12.44
4C	254	-3764.91	0.23	13.71	0.20	-63.75	0.68	0.94	-76.99	-61.94	-	-63.10	-76.11	-61.70	-75.16	-74.38
5A	260	-3619.63	0.20	76.81	63.30	-0.66	63.78	64.04	-13.90	1.16	63.10	-	-13.02	1.40	-12.07	-11.29
5B	262	-3589.66	0.22	89.83	76.31	12.36	76.79	77.05	-0.88	14.17	76.11	13.02	-	14.42	0.95	1.73
6	266	-3622.86	0.23	75.41	61.90	-2.06	62.37	62.64	-15.30	-0.25	61.70	-1.40	-14.42	-	-13.47	-12.69
7	272	-3591.85	0.23	88.88	75.36	11.41	75.84	76.10	-1.83	13.22	75.16	12.07	-0.95	13.47	-	0.78
8	278	-3593.64	0.21	88.10	74.59	10.63	75.06	75.33	-2.61	12.44	74.38	11.29	-1.73	12.69	-0.78	-

The Bayes factor cut-off value to select a more complex model was set to >100 (log_{10}BF>2). The bold row indicates the model that was finally selected for all further analyses. [a], number of free parameters in the model; [b], marginal likelihood of the model; [c], standard error of the marginal likelihood as determined by 1000 bootstrap replicates.

Supporting Table S2.5: Summary statistics for the relevant parameter estimates in BEAST (mtDNA)

		Quantiles								
Parameter		mean	mode	2.5%	10%	25%	median	75%	90%	97.5%
TMRCA Pongo [mya]	Posterior	3.50	3.35	2.40	2.72	3.05	3.45	3.89	4.32	4.86
	Prior	15.24	15.11	9.05	10.98	12.98	15.21	17.46	19.47	21.72
TMRCA Sumatra [mya]	Posterior	0.85	0.79	0.53	0.62	0.72	0.83	0.96	1.10	1.28
	Prior	12.64	11.99	6.63	8.38	10.26	12.49	14.90	17.06	19.41
TMRCA SQ/KE/TR/BA [mya]	Posterior	0.25	0.23	0.12	0.16	0.19	0.24	0.30	0.37	0.46
	Prior	9.13	8.07	3.91	5.39	6.90	8.81	11.13	13.29	15.71
TMRCA Borneo+BT [mya]	Posterior	2.09	1.93	1.32	1.56	1.78	2.05	2.36	2.68	3.07
	Prior	12.87	13.34	6.84	8.62	10.40	12.76	15.12	17.32	19.74
TMRCA Borneo [mya]	Posterior	0.18	0.16	0.09	0.11	0.13	0.17	0.21	0.26	0.33
	Prior	10.82	10.31	5.31	6.82	8.46	10.57	12.96	15.15	17.65

Supporting Table S2.6: Summary statistics for the relevant parameter estimates in BATWING (Y-chromosomal loci)

				Quantiles						
Parameter		mean	mode	2.5%	10%	25%	median	75%	90%	97.5%
TMRCA Pongo [yrs]	Posterior	1.68E5	1.05E5	5.57E4	7.45E4	9.88E4	1.38E5	2.00E5	2.91E5	4.62E5
	Prior	6.24E12	8.24E11	9.58E10	5.23E11	1.70E12	4.62E12	9.57E12	1.47E13	1.96E13
TMRCA Sumatra [yrs]	Posterior	4.19E3	3.23E3	1.68E3	2.26E3	2.91E3	3.82E3	5.08E3	6.51E3	8.86E3
	Prior	2.42E7	6.08E6	9.24E5	3.77E6	9.29E6	1.91E7	3.31E7	5.08E7	7.87E7
TMRCA Borneo [yrs]	Posterior	2.10E4	1.78E4	1.08E4	1.32E4	1.57E4	1.94E4	2.44E4	3.04E4	4.05E4
	Prior	2.47E7	8.90E6	9.47E5	3.90E6	9.66E6	1.96E7	3.36E7	5.14E7	7.89E7
N_e Pongo	Posterior	1213.21	1092.66	736.66	869.99	1000.43	1172.53	1382.17	1611.67	1923.53
	Prior	5.00E5	3.76E5	2.61E4	1.03E5	2.51E5	4.99E5	7.50E5	8.99E5	9.74E5
N_e Sumatra	Posterior	179.63	136.88	71.93	96.14	125.52	164.59	218.11	282.88	355.21
	Prior	4.98E5	1.48E5	2.50E4	9.79E4	2.46E5	4.98E5	7.47E5	8.97E5	9.74E5
N_e Borneo	Posterior	1082.95	963.92	647.38	760.05	883.95	1053.43	1242.23	1440.92	1714.21
	Prior	5.01E5	4.50E5	2.52E4	1.03E5	2.54E5	5.00E5	7.50E5	9.01E5	9.76E5

7.2. Chapter 3

Supporting Table S3.1: List of samples used for analyses

Sample ID	Type	Status[a]	Reliable[b]	Region[c]	Origin	Sex	HVRI[d]	STR[e]
BA1	Blood	R	yes	BA	Sidikalang, Dairi, North Sumatra	female		X
BA2	Blood	R	yes	BA	Sidikalang, Dairi, North Sumatra	female	wa1	X
BA3	Blood	R	yes	BA	Sidikalang, Dairi, North Sumatra	female		X
BA4	Blood	R	yes	BA	Perolihen, Pakpak Bharat, North Sumatra	male	wa1	X
BA5	Faeces	W		BA	Perolihen, Pakpak Bharat, North Sumatra	n/a	wa1	X
BA6	Hair	W		BA	Perolihen, Pakpak Bharat, North Sumatra	male		X
BA7	Hair	R	yes	BA	Salak, Pakpak Bharat, North Sumatra	female	wa1	X
BA8	Hair	R	yes	BA	Perolihen, Pakpak Bharat, North Sumatra	male	wa1	X
BA9	Hair	R	yes	BA	Singkohor, Aceh Singkil, Aceh	female		X
BA10	Hair	W		BA	Puncak Sidiangkat, Pakpak Bharat, North Sumatra	n/a	wa2	
BA11	Hair	W		BA	Siranggas, Pakpak Bharat, North Sumatra	n/a	wa1	
BA12	Hair	W		BA	Salak, Pakpak Bharat, North Sumatra	n/a	wa3	
BT1	Blood	R	yes	BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	male	bt2	X
BT2	Blood	R	yes	BT	Tukka, Tapanuli Tengah, North Sumatra	female	bt4	X
BT3	Blood	R	yes	BT	Pinang Sori, Tapanuli Tengah, North Sumatra	male	bt4	X
BT4	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	male	bt1	X
BT5	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	n/a	bt3	
BT6	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	female	bt3	X
BT7	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	male	bt4	
BT8	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	female	bt3	X
BT9	Faeces	W		BT	West Batang Toru forest, Tapanuli Tengah, North Sumatra	female	bt4	X
BT10	Hair	R	yes	BT	Marancar, Tapanuli Selatan, North Sumatra	male	bt4	
BT11	Hair	W		BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	n/a	bt4	
BT12	Hair	W		BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	n/a	bt4	
BT13	Hair	W		BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	n/a	bt4	
BT14	Hair	W		BT	West Batang Toru forest, Tapanuli Selatan, North Sumatra	n/a	bt4	
BT15	Hair	W		BT	Rambasiashur, Tapanuli Selatan, North Sumatra	n/a	bt1	
BT16	Hair	W		BT	Rambasiashur, Tapanuli Selatan, North Sumatra	n/a	bt1	
BT17	Hair	W		BT	Rambasiashur, Tapanuli Selatan, North Sumatra	n/a	bt1	X
BT18	Hair	W		BT	Sitandiang, Tapanuli Selatan, North Sumatra	n/a	bt4	

Sample ID	Type	Status[a]	Reliable[b]	Region[c]	Origin	Sex	HVRI[d]	STR[e]
CL1	Blood	R	yes	CL	Ladang Lawe Bengkari, Aceh Tenggara, Aceh	male	wa1	X
CL2	Blood	R	yes	CL	Lawe Sikap, Aceh Tenggara, Aceh	male	wa1	X
CL3	Blood	R	yes	CL	Lawe Sikap, Aceh Tenggara, Aceh	male	wa1	X
CL4	Blood	R	yes	CL	Lawe Sikap, Aceh Tenggara, Aceh	male	wa1	X
CL5	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	
CL6	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female	wa1	
CL7	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female	wa5	
CL8	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male		X
CL9	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female		X
CL10	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	
CL11	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	
CL12	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female		X
CL13	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	X
CL14	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa4	X
CL15	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	X
CL16	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	female		X
CL17	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male	wa1	X
CL18	Faeces	W		CL	Ketambe, West Side of Alas River, Aceh Tenggara, Aceh	male		X
CL19	Hair	R	yes	CL	Lawe Pupus, Aceh Tenggara, Aceh	male		X
CL20	Hair	R	yes	CL	Pindie, Gayo Lues, Aceh	female	wa1	X
LK1	Blood	R	yes	LK	Tangkahan, Langkat, North Sumatra	male	lk1	X
LK2	Blood	R	yes	LK	Langkat, North Sumatra	male	lk2	X
LK3	Blood	R	yes	LK	Pinding, Gayo Lues, Aceh	female	lk2	X
LK4	Blood	R	yes	LK	Sampan Getek, Langkat, North Sumatra	female	lk1	X
LK5	Blood	R	yes	LK	Sampan Getek, Langkat, North Sumatra	female	lk1	X
LK6	Blood	R	yes	LK	Blangkejeren, Gayo Lues, Aceh	female	lk2	X
LK7	Blood	R	yes	LK	Rikit Gaib, Gayo Lues, Aceh	male	lk2	X
LK8	Blood	R	yes	LK	Tenggulun, Aceh Tamiang, Aceh	male	lk7	X
LK9	Blood	R	yes	LK	Tamiang Hulu, Aceh Tamiang, Aceh	male	lk7	X
LK10	Blood	R	yes	LK	Kejuruan Muda, Aceh Tamiang, Aceh	female	lk7	X
LK11	Faeces	W		LK	Ketambe, East Side of Alas River, Aceh Tenggara, Aceh	female	lk2	X
LK12	Faeces	W		LK	Ketambe, East Side of Alas River, Aceh Tenggara, Aceh	female	lk2	X
LK13	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk3	
LK14	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	male	lk1	X
LK15	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk1	X
LK16	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	male	lk1	X

Sample ID	Type	Status[a]	Reliable[b]	Region[c]	Origin	Sex	HVRI[d]	STR[e]
LK17	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk1	
LK18	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk1	
LK19	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female	lk1	X
LK20	Faeces	W		LK	Sikundur, Langkat, North Sumatra	female	lk4	X
LK21	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	female		X
LK22	Faeces	W		LK	Sampan Getek, Langkat, North Sumatra	male		X
LK23	Faeces	W		LK	Aras Napal, Langkat, North Sumatra	n/a	lk7	
LK24	Faeces	W		LK	Aras Napal, Langkat, North Sumatra	n/a	lk4	
LK25	Faeces	W		LK	Aras Napal, Langkat, North Sumatra	n/a	lk4	X
LK26	Faeces	W		LK	Aras Napal, Langkat, North Sumatra	female	lk7	X
LK27	Hair	R	yes	LK	Kota Panjang, Gayo Lues, Aceh	female	lk5	X
LK28	Hair	R	yes	LK	Pinding, Gayo Lues, Aceh	male		X
LK29	Hair	R	yes	LK	Barak Gajah, Langkat, North Sumatra	female	lk6	X
NA1	Blood	R	yes	NA	Langsa, Aceh	male	na2	
NA2	Blood	R	yes	NA	Pondok Baru, Aceh Tengah, Aceh	male	na1	X
NA3	Blood	R	yes	NA	Lhokseumawe, Aceh	male	na1	X
NA4	Blood	R	yes	NA	Gayo Lues, Aceh	male	na1	X
NA5	Blood	R	yes	NA	Langsa, Aceh	male	na1	X
NA6	Blood	R	yes	NA	Aceh Besar, Aceh	male		X
NA7	Blood	R	yes	NA	Blang Jerango, Agara, Gayo Lues, Aceh	male	na1	X
NA8	Blood	R	yes	NA	Indra Makmu, Aceh Timur, Aceh	male	na3	X
NA9	Blood	R	yes	NA	Indra Makmu, Aceh Timur, Aceh	male	na3	X
NA10	Hair	R	yes	NA	Takengon, Aceh Tengah, Aceh	female	na1	X
NA11	Hair	R	yes	NA	Takengon, Aceh Tengah, Aceh	male	na1	X
WL1	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	male	wa6	
WL2	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	female	wa1	
WL3	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	female	wa7	
WL4	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	male	wa8	
WL5	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	female	wa6	
WL6	Blood	R	yes	WL	Aceh Selatan, near Suaq Balimbing, Aceh	female		X
WL7	Blood	R	yes	WL	Singkil, Aceh Singkil, Aceh	female	wa1	X
WL8	Blood	R	yes	WL	Gunung Merutung, Aceh Selatan, Aceh	male	wa1	X
WL9	Blood	R	yes	WL	Cot Siaumantouk, Aceh Selatan, Aceh	female	wa9	X
WL10	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female		X
WL11	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL12	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X

Sample ID	Type	Status[a]	Reliable[b]	Region[c]	Origin	Sex	HVRI[d]	STR[e]
WL13	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL14	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female	wa1	X
WL15	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL16	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	X
WL17	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL18	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL19	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female	wa1	X
WL20	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female		X
WL21	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa6	X
WL22	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL23	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female	wa6	
WL24	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL25	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL26	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female	wa6	X
WL27	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female		X
WL28	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male		X
WL29	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa6	
WL30	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	female		X
WL31	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL32	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL33	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL34	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL35	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL36	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL37	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL38	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa1	
WL39	Faeces	W		WL	Suaq Balimbing, Aceh Selatan, Aceh	male	wa6	
WL40	Hair	R	yes	WL	Singkil, Aceh Singkil, Aceh	male	wa1	
WL41	Hair	R	yes	WL	Singkil, Aceh Singkil, Aceh	male	wa1	
TR1	Blood	R	yes	TR	Trangon, Gayo Lues, Aceh	female		X
TR2	Blood	R	yes	TR	Aceh Barat Daya, Aceh	male	wa10	X
TR3	Blood	R	yes	TR	Aceh Barat, Aceh	male	wa11	X
TR4	Blood	R	yes	TR	Aluebillie, Aceh Nagan Raya, Aceh	male	wa1	X
TR5	Blood	R	yes	TR	Aceh Barat Daya, Aceh	female	wa12	X
TR6	Blood	R	yes	TR	Trangon, Gayo Lues, Aceh	male	wa13	X
TR7	Hair	R	yes	TR	Meulaboh, Aceh Barat, Aceh	female		X

Sample ID	Type	Status[a]	Reliable[b]	Region[c]	Origin	Sex	HVRI[d]	STR[e]
TR8	Hair	R	yes	TR	Meulaboh, Aceh Barat, Aceh	female	wa9	X
TR9	Hair	R	yes	TR	Trangon, Gayo Lues, Aceh	male	wa1	X
UNK1	Blood	R	no	UNK	Sumatra	male	bt1	X
UNK2	Blood	R	no	UNK	Gayo Lues, Aceh	male	na1	X
UNK3	Blood	R	no	UNK	Northern Aceh	male	na1	X
UNK4	Blood	R	no	UNK	Sumatra	male	na1	X
UNK5	Blood	R	no	UNK	Sumatra	female	na2	X
UNK6	Blood	R	no	UNK	Sumatra	female	na1	X
UNK7	Blood	R	no	UNK	Northern Aceh	male	na1	X
UNK8	Blood	R	no	UNK	Aceh	female	wa14	X
UNK9	Blood	R	no	UNK	Aceh	male	wa1	X
UNK10	Blood	R	no	UNK	Aceh	female		X
UNK11	Blood	R	no	UNK	Aceh	male	wa1	X
UNK12	Blood	R	no	UNK	Aceh	female	wa1	X
UNK13	Blood	R	no	UNK	Aceh	female		X

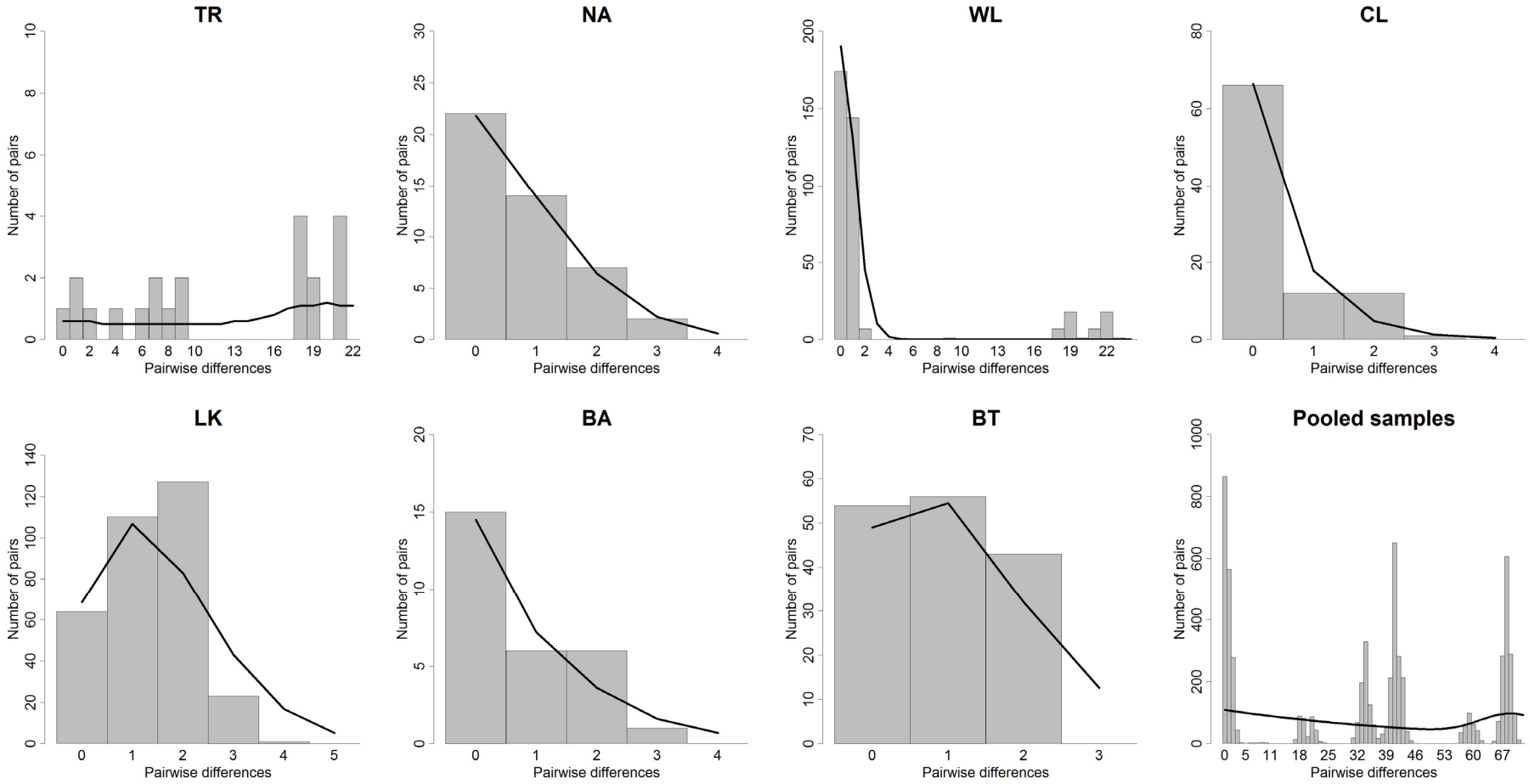
[a] sample status, W=sampled in the wild, R=rehabilitant individual; [b] provenance reliability of samples from rehabilitant individuals; [c] sampling region used for analyses (see section 3.3); [d] mitochondrial HVRI haplotype; [e] autosomal microsatellite data included in analyses.

Supporting Table S3.2: Summary statistics for the seven sampling regions and pooled samples

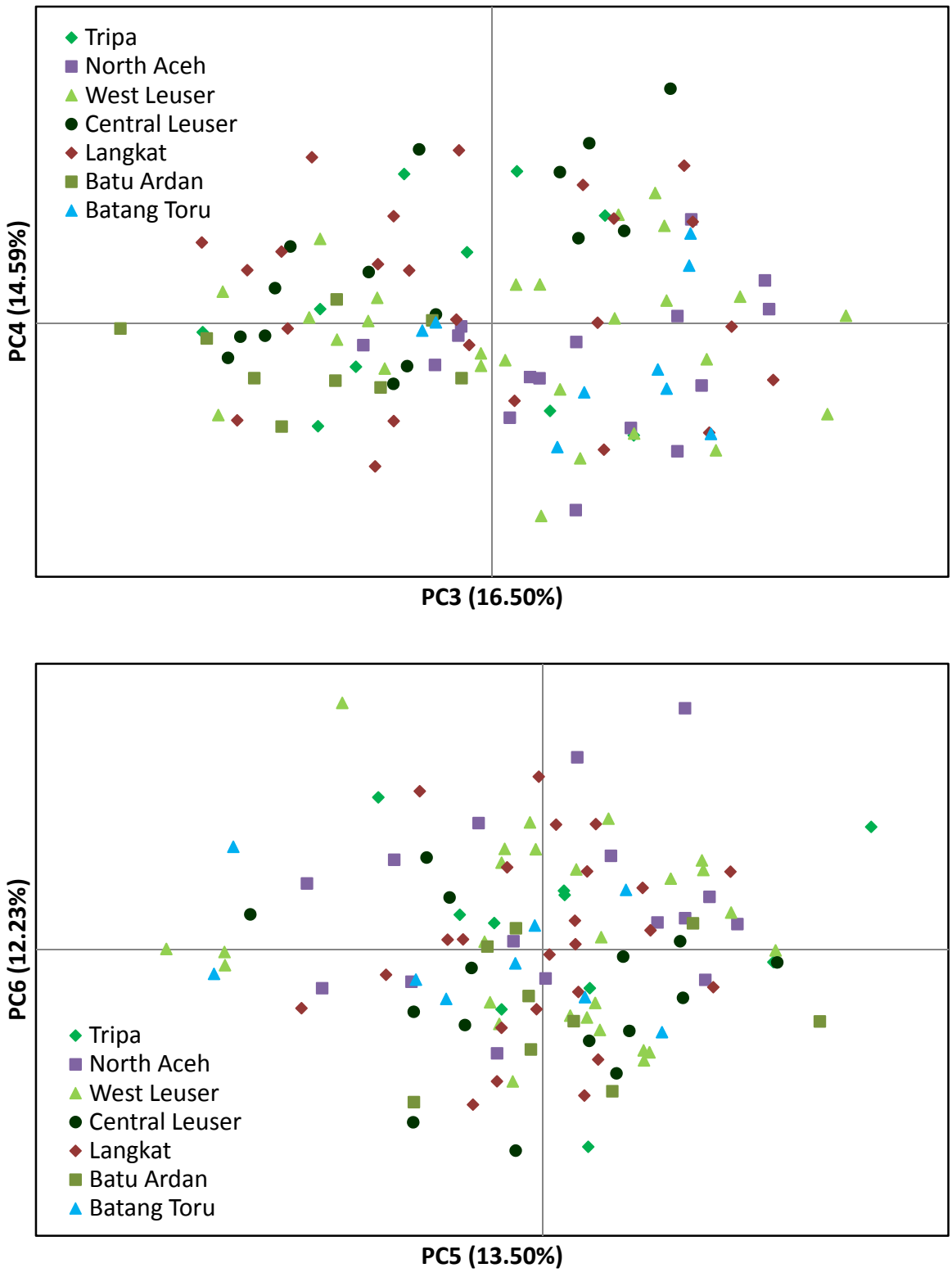
	TR	NA	WL	CL	LK	BA	BT	Pooled
HVRI								
θ_π	12.78	0.79	3.78	0.44	1.40	0.78	0.96	68.40
θ_S	8.98	1.06	5.91	0.94	1.57	1.16	0.87	18.74
θ_L mode	10.09	1.05	3.88	1.12	1.91	1.16	1.00	19.03
θ_L 95%-HPD[a]	4.33–30.64	0.20–4.13	1.92–7.89	0.19–3.53	0.62–4.46	0.23–4.91	0.17–2.91	14.32–26.35
$N_e(\theta_\pi)$	6,808	419	2,012	236	747	417	512	36,432
$N_e(\theta_S)$	4,783	565	3,148	502	838	616	465	9,983
$N_e(\theta_L)$	5,448	566	2,095	604	1,032	627	541	10,271
Tajima's D[b]	0.76	-1.03	-1.79*	-1.67*	-0.43	-1.45	0.18	2.28
Fu's F_S[c]	0.44	-0.05	2.97	-0.76	-1.96	-0.30	-0.29	15.30
SSD[d]	0.07	0.00	0.02	0.01	0.02	0.01	0.01	0.05*
RI[e]	0.16	0.07	0.14	0.37	0.13	0.14	0.09	0.03***
Microsatellites								
θ_H	1.68	1.60	1.61	1.56	1.66	1.57	1.63	1.70
θ_L mode	3.95	4.04	4.11	4.47	4.28	4.06	3.40	4.27
θ_L 95%-HPD[a]	2.94–4.77	3.97–4.67	4.00–4.53	4.30–4.73	4.22–4.69	3.93–5.14	2.63–4.30	4.05–4.55
$N_e(\theta_H)$	4,197	4,009	4,023	3,901	4,162	3,929	4,087	4,257
$N_e(\theta_L)$	9,870	10,098	10,286	11,174	10,696	10,153	8,502	10,685

* $p<0.05$; ** $p<0.01$; *** $p<0.001$

[a] 95%-highest posterior density interval; [b] Tajima's D test (Tajima 1989), significantly positive value indicate population contraction, negative values population expansion; [c] Fu's F_S test (Fu 1997), significantly negative values indicate population expansion; [d] sum of squared deviations from the expectations under a model of population expansion (Schneider & Excoffier 1999), significance indicates stationarity; [e] raggedness index (Harpending et al. 1998), significance indicates stationarity.



Supporting Figure S3.3: Pairwise mismatch distributions for the HVRI haplotypes of seven sampling regions and all pooled samples. The gray bars indicate the observed number of sequence pairs for each category of pairwise sequence differences. The black line indicates the expectation under a model of spatial expansion.



Supporting Figure S3.4: Additional principal components for the microsatellite genotypes of the seven sampling regions.

7.3. Chapter 4

Supporting Table S4.1: Primers used for amplification and sequencing of four autosomal and one X-chromosomal region

Primer Name	Primer Type	Annealing Temp. (PCR/Sequencing)	Sequence (5'-3')
Chr2a_Region17_F	PCR / Sequencing primer	64°C / 53°C	AGTGCCCCGACACAAGTGATACAG
Chr2a_Region17_R	PCR / Sequencing primer	64°C /53°C	GAGCAGGGCTTAGGCAAGGAGA
Chr2a_Region17_seq1	Sequencing primer	53°C	GTTTTGAAGCCATTAAGTTGCTGAT
Chr2a_Region17_seq2	Sequencing primer	53°C	GGTGGAAACATTTTCAAAACTCAGA
Chr9_Region16_F	PCR / Sequencing primer	64°C / 53°C	TTCATATGCAGGGCAAGAGAACAAG
Chr9_Region16_R	PCR / Sequencing primer	64°C /53°C	CCCTGGTCATCATGCCTGCTATTAT
Chr9_Region16_seq1	Sequencing primer	53°C	AAGTTCACAGCCTTCCTCAAGAG
Chr12_Region1_F	PCR / Sequencing primer	64°C / 53°C	ATCCAAATGGCCAAACTCACCT
Chr12_Region1_R	PCR / Sequencing primer	64°C /53°C	GCAACCCACATGCTCATCAATAG
Chr12_Region1_seq1	Sequencing primer	53°C	CCAGGGAGAGCCAGGGAACA
Chr19_R7_F	PCR / Sequencing primer	64°C / 53°C	GGAGGGTTGATGACGTTTACTTACA
Chr19_R7_R	PCR / Sequencing primer	64°C /53°C	TGACACATGATTGATGCCACTCTC
Chr19_R7_seq1	Sequencing primer	53°C	AGGATACAAGCCCTATTTTGCTGAA
Xq13.3_2_F	PCR / Sequencing primer	62°C / 53°C	CTCAGTAACTTGGCGAAACCTCAT
Xq13.3_2_R	PCR / Sequencing primer	62°C / 53°C	GCCCCCAACAGACTCCAGTGT
Xq13.3_2_seq1	Sequencing primer	53°C	TGCAGCAACTAACAGCATTCA
Xq13.3_3_F	PCR / Sequencing primer	62°C / 53°C	TAAGTGGGAGCTGAATGATAAGAAC
Xq13.3_3_R	PCR / Sequencing primer	62°C / 53°C	GACAGGGAAGATTGAGAGTGAAGAT
Xq13.3_3_seq1	Sequencing primer	53°C	TCCCATGAAACACTCTCCTAAACA
Xq13.3_4_F	PCR / Sequencing primer	62°C / 53°C	CCCCTCTGAACCCTGCTCCTA
Xq13.3_4_R	PCR / Sequencing primer	62°C / 53°C	CCCTGGACTTGTAGAAAAATCTGCT
Xq13.3_4_seq1	Sequencing primer	53°C	ATAATCATGTTCTTTGGAAGACCTG
Xq13.3_5_F	PCR / Sequencing primer	62°C / 53°C	AAATCTTCTTAACTGTTGGGCACTT
Xq13.3_5_R	PCR / Sequencing primer	62°C / 53°C	TTAACGTTAACGCCATCAGTCC
Xq13.3_5_seq1	Sequencing primer	53°C	GGCAATTGGGAAAGGATACTCA
Xq13.3_5_seq2	Sequencing primer	53°C	AGCCAGAGTCTTGGTTTGTCTCC

The naming of the regions correspond to the names used in Fischer et al. (2006).

Supporting Table S4.2: Parameterization and parameter prior distributions for all eight tested demographic models

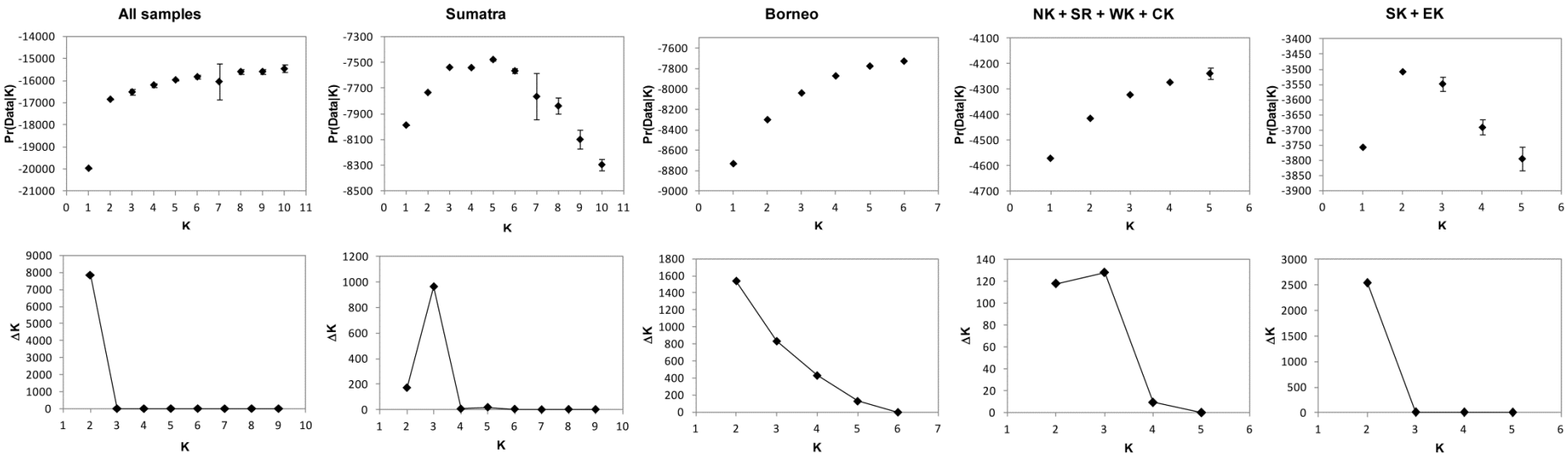
Parameter[a]	Prior distribution	I2[b]	IM2[c]	IM2-GR[d]	IM2-BN-GR[e]	IM10-DECST[f]	IM10-BN-DECST[g]	IM10-BN-DECSU[h]	IM10-BN-DECALL[i]
N_NOW_BO	log uniform	3, 5	3, 5	3, 5	3, 5	2,5	2,5	2,5	2,5
N_NOW_SU	log uniform	3, 5	3, 5	3, 5	3, 5				
N_NOW_NT	log uniform					2,5	2,5	2,5	2,5
N_NOW_ST	log uniform					2,4	2,4	2,4	2,4
N_BN_BO	log uniform				2, 5		2, 5	2, 5	2, 5
N_BN_SU	log uniform				2, 5				
N_STRUC_BO	log uniform								3, 5
N_STRUC_NT	log uniform							3, 5	3, 5
N_ANC_BO	log uniform			3, 6	3, 6	3, 6	3, 6	3, 6	3, 6
N_ANC_SU	log uniform			3, 6	3, 6				
N_ANC_NT	log uniform					3, 6	3, 6	3, 6	3, 6
N_ANC_ST	log uniform					3, 6	3, 6	3, 6	3, 6
N_ANC_PO	log uniform	3, 6	3, 6	3, 6	3, 6				
T_SPLIT	log uniform	4, 5	4, 5	4, 5	4, 5				
T_SPLIT_BO	log uniform					4.2 ,4.8	4.2 ,4.8	4.2 ,4.8	4.2 ,4.8
T_SPLIT_NT	log uniform					4.8, 5.2	4.8, 5.2	4.8, 5.2	4.8, 5.2
T_MIGSTOP	log uniform		2, 5	2, 5	2, 5	2, 4.2	2, 4.2	2, 4.2	2, 4.2
T_BN_BO	log uniform				2, 5	2, 4.2	2, 4.2	2, 4.2	2, 4.2
T_BN_SU	log uniform				2, 5				
T_STRUC_BO	log uniform					2.5, 4.2	2.5, 4.2	2.5, 4.2	2.5, 4.2
T_STRUC_NT	log uniform					3.5, 4.8	3.5, 4.8	3.5, 4.8	3.5, 4.8
T_DEC_BO	log uniform								0, 2.5
T_DEC_NT	log uniform							1, 3.5	1, 3.5
T_DEC_ST	log uniform					1, 3.5	1, 3.5	1, 3.5	1, 3.5
m_BO-SU	log uniform		-6, -2	-6, -2	-6, -2				
m_SU-BO	log uniform		-6, -2	-6, -2	-6, -2				
m_BO-ST	log uniform					-6, -2	-6, -2	-6, -2	-6, -2
m_ST-BO	log uniform					-6, -2	-6, -2	-6, -2	-6, -2
m_NT-ST	log uniform					-6, -2	-6, -2	-6, -2	-6, -2
m_ST-NT	log uniform					-6, -2	-6, -2	-6, -2	-6, -2
MALEMIG	uniform		0, 1	0, 1	0, 1	0, 1	0, 1	0, 1	0, 1

Parameter[a]	Prior distribution	I2[b]	IM2[c]	IM2-GR[d]	IM2-BN-GR[e]	IM10-DECST[f]	IM10-BN-DECST[g]	IM10-BN-DECSU[h]	IM10-BN-DECALL[i]
ALPHA_STR	uniform	8, 15	8, 15	8, 15	8, 15	8, 15	8, 15	8, 15	8, 15
ALPHA_Y_STR	uniform	8, 15	8, 15	8, 15	8, 15	8, 15	8, 15	8, 15	8, 15
STR_MUT	log uniform	-5, -3	-5, -3	-5, -3	-5, -3	-5, -3	-5, -3	-5, -3	-5, -3
Y_STR_MUT	normal	2.0×10^{-3}, 1.0×10^{-3}	2.0×10^{-3}, 1.0×10^{-3}	2.0×10^{-3}, 1.0×10^{-3}	2.0×10^{-3}, 1.0×10^{-3}	2.0×10^{-3}, 1.0×10^{-3}	2.0×10^{-3}, 1.0×10^{-3}	2.0×10^{-3}, 1.0×10^{-3}	2.0×10^{-3}, 1.0×10^{-3}
Chr2a_MUT	normal	1.64×10^{-8}, 2.98×10^{-9}	1.64×10^{-8}, 2.98×10^{-9}	1.64×10^{-8}, 2.98×10^{-9}	1.64×10^{-8}, 2.98×10^{-9}	1.64×10^{-8}, 2.98×10^{-9}	1.64×10^{-8}, 2.98×10^{-9}	1.64×10^{-8}, 2.98×10^{-9}	1.64×10^{-8}, 2.98×10^{-9}
Chr9_MUT	normal	3.02×10^{-8}, 6.58×10^{-9}	3.02×10^{-8}, 6.58×10^{-9}	3.02×10^{-8}, 6.58×10^{-9}	3.02×10^{-8}, 6.58×10^{-9}	3.02×10^{-8}, 6.58×10^{-9}	3.02×10^{-8}, 6.58×10^{-9}	3.02×10^{-8}, 6.58×10^{-9}	3.02×10^{-8}, 6.58×10^{-9}
Chr12_MUT	normal	1.85×10^{-8}, 3.46×10^{-9}	1.85×10^{-8}, 3.46×10^{-9}	1.85×10^{-8}, 3.46×10^{-9}	1.85×10^{-8}, 3.46×10^{-9}	1.85×10^{-8}, 3.46×10^{-9}	1.85×10^{-8}, 3.46×10^{-9}	1.85×10^{-8}, 3.46×10^{-9}	1.85×10^{-8}, 3.46×10^{-9}
Chr19_MUT	normal	2.21×10^{-8}, 5.25×10^{-9}	2.21×10^{-8}, 5.25×10^{-9}	2.21×10^{-8}, 5.25×10^{-9}	2.21×10^{-8}, 5.25×10^{-9}	2.21×10^{-8}, 5.25×10^{-9}	2.21×10^{-8}, 5.25×10^{-9}	2.21×10^{-8}, 5.25×10^{-9}	2.21×10^{-8}, 5.25×10^{-9}
Xq13.3_MUT	normal	1.96×10^{-8}, 2.60×10^{-9}	1.96×10^{-8}, 2.60×10^{-9}	1.96×10^{-8}, 2.60×10^{-9}	1.96×10^{-8}, 2.60×10^{-9}	1.96×10^{-8}, 2.60×10^{-9}	1.96×10^{-8}, 2.60×10^{-9}	1.96×10^{-8}, 2.60×10^{-9}	1.96×10^{-8}, 2.60×10^{-9}
MTDNA_MUT	normal	2.80×10^{-7}, 4.80×10^{-8}	2.80×10^{-7}, 4.80×10^{-8}	2.80×10^{-7}, 4.80×10^{-8}	2.80×10^{-7}, 4.80×10^{-8}	2.80×10^{-7}, 4.80×10^{-8}	2.80×10^{-7}, 4.80×10^{-8}	2.80×10^{-7}, 4.80×10^{-8}	2.80×10^{-7}, 4.80×10^{-8}

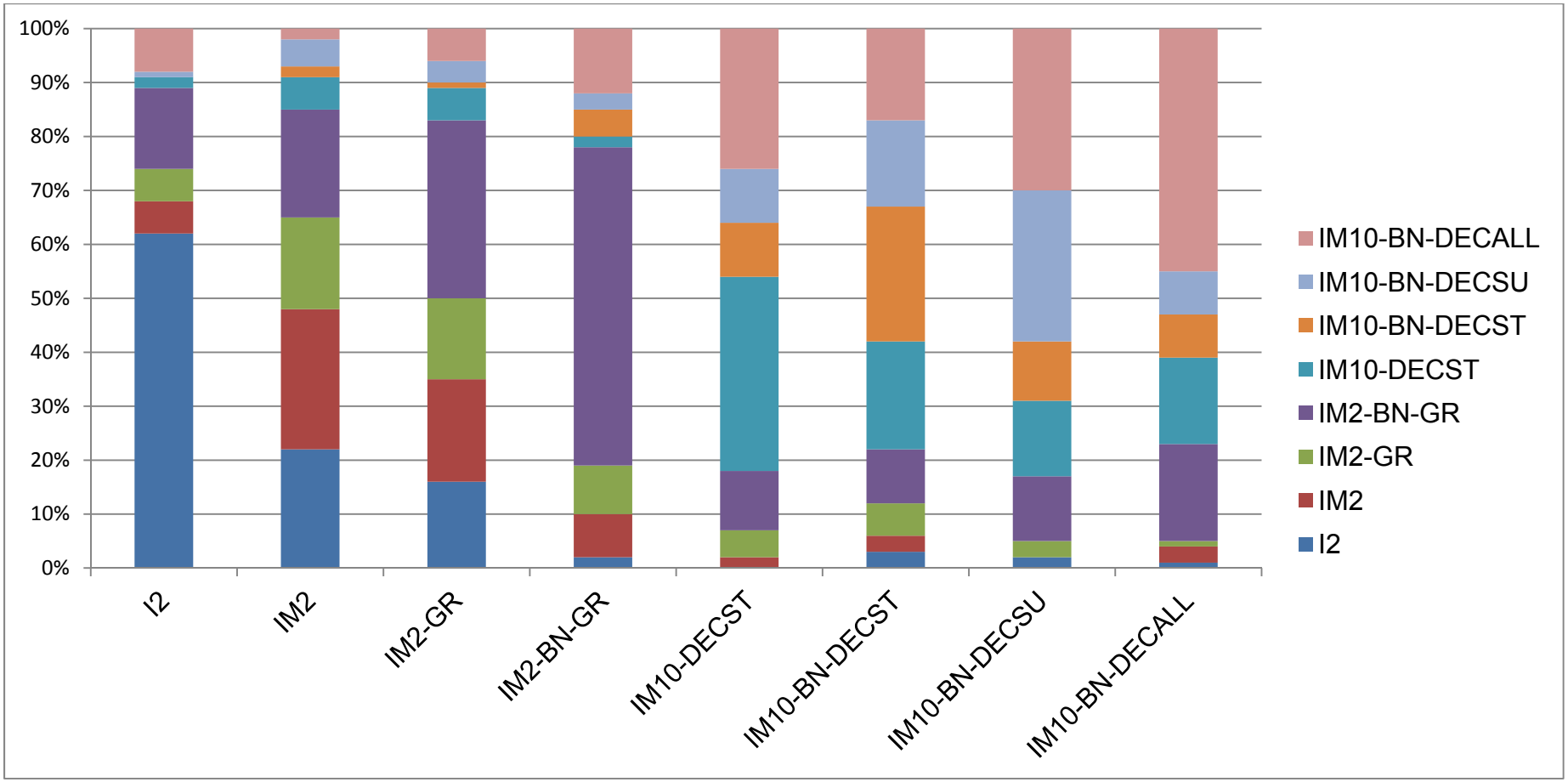
[a], BO = Borneo, SU = Sumatra, NT = Sumatra north of Lake Toba, ST = Sumatra south of Lake Toba, N_NOW = current effective population size, N_BN = effective population size during population bottleneck, N_STRUC = effective population size before recent decline, N_ANC = ancestral effective population size, T_SPLIT = population split time, T_MIGSTOP = time since migration between Borneo and Sumatra stopped, T_BN = time since population bottleneck, T_STRUC = time since establishment of population structure, T_DEC = time since population decline, m = migration rate per individual per generation, ALPHA = shape parameter of gamma distribution of mutation rate, MUT = mean mutation rate per locus/site per generation; [b], isolation model with two populations; [c], isolation-with-migration with two populations; [d], isolation-with-migration model with two populations and exponential growth; [e], isolation-with-migration model with two populations and bottleneck followed by exponential growth; [f], isolation-with-migration model with 10 populations and recent decline in population south of Lake Toba; [g], isolation-with-migration model with 10 populations, bottleneck on Borneo and recent decline in population south of Lake Toba; [h], isolation-with-migration model with 10 populations, bottleneck on Borneo and recent decline in all Sumatran populations; [i], isolation-with-migration model with 10 populations, bottleneck on Borneo and recent decline in all populations.

Supporting Table S4.3: Summary statistics used for the Approximate Bayesian Computation

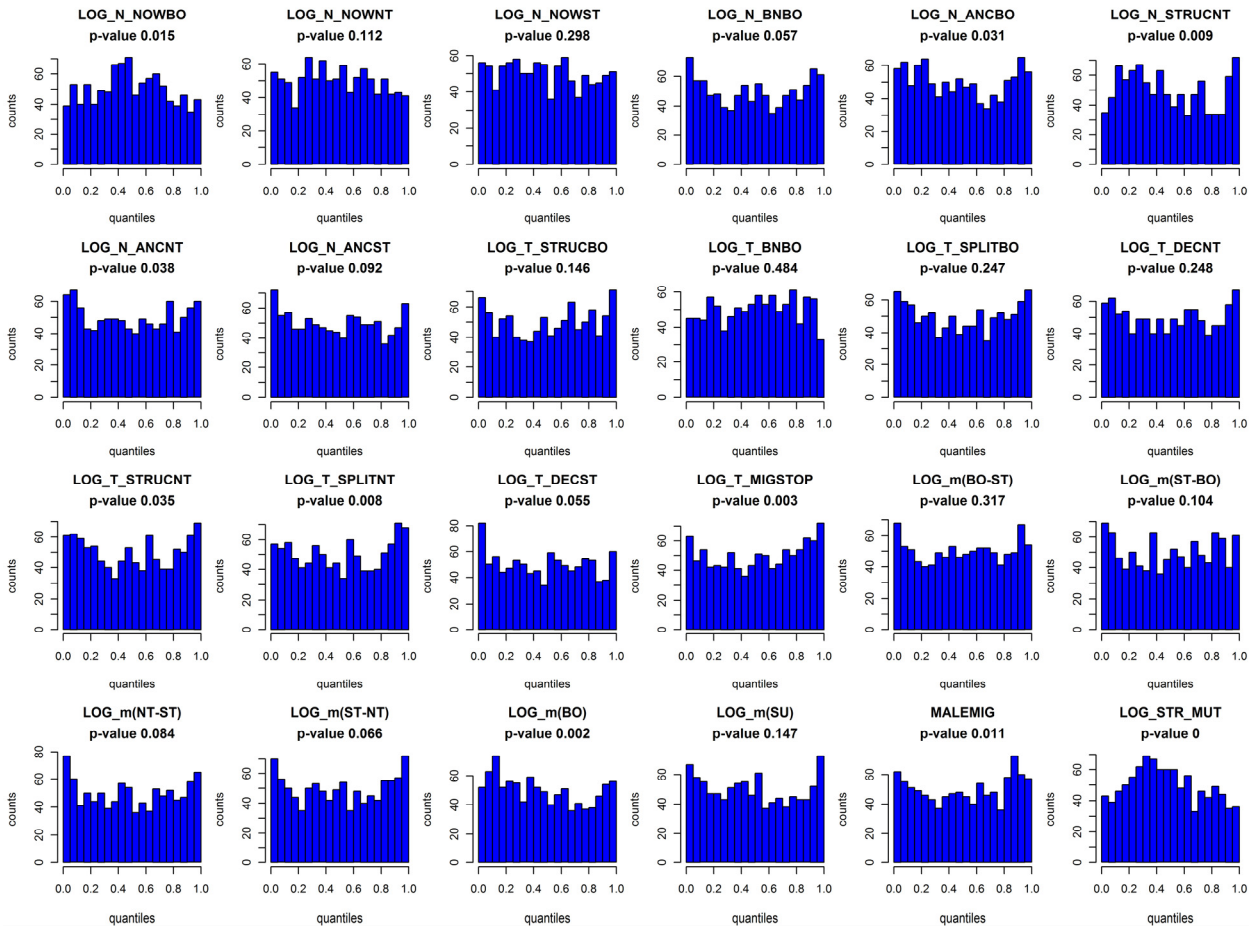
Summary statistic	Data type (number of data sets)	Number of statistics	Explanation
S_X	DNA (3)	6	Number of segregating sites per population
prS_X	DNA (3)	6	Number of private segregating sites per population
S_{tot}	DNA (3)	3	Total number of segregating sites over all populations
D_X	DNA (3)	6	Tajima's D (Tajima 1989), calculated for each population.
π_X	DNA (3)	6	Average number of pairwise sequence differences within each population
Φ_{ST_XY}	DNA (3)	3	Differentiation index between all pairs of populations, calculated as Φ_{ST} (Excoffier et al. 1992).
π_{XY}	DNA (3)	3	Mean number of sequence differences between all pairs of populations
K_X	Microsatellites (2)	4	Mean number of alleles over all loci per population (autosomal) or number of haplotypes per population (Y-chromosomal)
K_{tot}	Microsatellites (2)	2	Mean over all loci of the total number of alleles in all populations (autosomal) or total number of haplotypes in all populations (Y-chromosomal)
H_X	Microsatellites (2)	4	Mean heterozygosity over all loci per population
H_{tot}	Microsatellites (2)	2	Mean over all loci of the total heterozygosity in all populations
GW_X	Microsatellites (2)	4	Mean Garza-Williamson index (Garza & Williamson 2001) over all loci per population ($GW_X = K_X/(R_X+1)$)
GW_{tot}	Microsatellites (2)	2	Mean Garza-Williamson index (Garza & Williamson 2001) over all loci over all populations ($GW_X=K_{tot}/(R_{tot}+1)$)
NGW_X	Microsatellites (2)	4	Mean modified Garza-Williamson index (Garza & Williamson 2001) over all loci per population ($NGW_X=K_X/(R_{tot}+1)$)
R_X	Microsatellites (2)	4	Mean allelic size range over all loci per population
R_{tot}	Microsatellites (2)	2	Mean over all loci of the total allelic size range in all populations
F_{IS}	Microsatellites (1)	1	Mean of the global inbreeding coefficient over all loci
F_{ST_XY}	Microsatellites (2)	2	Differentiation index between all pairs of populations, calculated as θ_W (Weir & Cockerham 1984).
π_{XY}	Microsatellites (2)	2	Mean number of allelic differences between all pairs of populations
$(\delta\mu)^2x_Y$	Microsatellites (2)	2	Square difference of mean within population repeat size between all pairs of populations (Goldstein et al. 1995)



Supporting Figure S4.4: Pr(Data|K) and deltaK statistics for all STRUCTURE runs. The population structure analysis incorporated multiple levels of hierarchical structure, starting with all samples and subsequently reducing the data set to only samples assigned to the same cluster in the previous analysis.



Supporting Figure S4.5: Cross validation of the model selection procedure. Every bar represents the proportions of assignments of 100 pseudo-observed datasets simulated under a given demographic model to one of the eight tested demographic models.



Supporting Figure S4.6: Cross validation of the parameter estimation. For every model parameter, we generated 1,000 random datasets under the selected demographic model (IM10-BN-DECSU). We then performed the standard parameter estimation procedure with each dataset. The histograms represent the number of times the known parameter values fall into each 10%-quantile of the estimated posterior distribution. For unbiased parameter estimates, the expectation is a uniform distribution over the entire prior space. A concentration of data points at the borders indicate too narrow posterior estimates, while a concentration of data points at the center points toward too conservative posterior estimates. The p-value of the Kolmogorov-Smirnov test is given above each histogram.

8. Co-authored Publications

Arora N, van Noordwijk MA, Ackermann C, Willems EP, Nater A, et al. (2012) Parentage-based pedigree reconstruction reveals female matrilineal clusters and male-biased dispersal in nongregarious Asian great apes, the Bornean orang-utans (Pongo pygmaeus). *Molecular Ecology* **21**, 3352-3362.

Nietlisbach P, Arora N, Nater A, Goossens B, van Schaik CP, Krützen M (2012) Heavily male-biased long-distance dispersal of orang-utans (genus: *Pongo*), as revealed by Y-chromosomal and mitochondrial genetic markers. *Molecular Ecology* **21**, 3173-3186.

Arora N, Nater A, van Schaik CP, et al. (2010) Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (Pongo pygmaeus). *Proceedings of the National Academy of Sciences* **107**, 21376-21381.

Nietlisbach P, Nater A, Greminger MP, Arora N, Krützen M (2010) A multiplex-system to target 16 male-specific and 15 autosomal genetic markers for orang-utans (genus: *Pongo*). *Conservation Genetics Resources* **2**, 153-158.

MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 3352–3362 doi: 10.1111/j.1365-294X.2012.05608.x

Parentage-based pedigree reconstruction reveals female matrilineal clusters and male-biased dispersal in nongregarious Asian great apes, the Bornean orang-utans (*Pongo pygmaeus*)

N. ARORA*, M. A. VAN NOORDWIJK*, C. ACKERMANN*, E. P. WILLEMS*, A. NATER*, M. GREMINGER*, P. NIETLISBACH*†, L. P. DUNKEL*, S. S. UTAMI ATMOKO‡, JOKO PAMUNGKAS§, DYAH PERWITASARI-FARAJALLAH§, C. P. VAN SCHAIK* and M. KRÜTZEN*

*Anthropological Institute & Museum, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland, †Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland, ‡Fakultas Biologi, Universitas Nasional, Jakarta 12510, Indonesia, §Indonesian Primate Centre, Bogor Agricultural University, Bogor 16680, Indonesia

Abstract

Philopatry and sex-biased dispersal have a strong influence on population genetic structure, so the study of species dispersal patterns and evolutionary mechanisms shaping them are of great interest. Particularly nongregarious mammalian species present an underexplored field of study: despite their lower levels of sociality compared to group-living species, interactions among individuals do occur, providing opportunities for cryptic kin selection. Among the least gregarious primates are orang-utans (genus: *Pongo*), in which preferential associations among females have nevertheless been observed, but for which the presence of kin structures was so far unresolved because of the equivocal results of previous genetic studies. To clarify relatedness and dispersal patterns in orang-utans, we examined the largest longitudinal set of individuals with combined genetic, spatial and behavioural data. We found that males had significantly higher mitochondrial DNA (mtDNA) variation and more unique haplotypes, thus underscoring their different maternal ancestries compared to females. Moreover, pedigree reconstruction based on 24 highly polymorphic microsatellite markers and mtDNA haplotypes demonstrated the presence of three matrilineal clusters of generally highly related females with substantially overlapping ranges. In orang-utans and possibly other nongregarious species, comparing average biparental relatedness (*r*) of males and females to infer sex-biased dispersal is extremely problematic. This is because the opportunistic sampling regime frequently employed in nongregarious species, combined with overlapping space use of distinct matrilineal clusters, leads to a strong downward bias when mtDNA lineage membership is ignored. Thus, in nongregarious species, correct inferences of dispersal can only be achieved by combining several genetic approaches with detailed spatial information.

Keywords: kin structure, matrilineal cluster, nongroup-living species, relatedness

Received 20 December 2011; revision accepted 19 March 2012

Introduction

Sex-biased natal dispersal, whereby one sex displays a greater tendency to leave or travel longer distances away from the natal area before breeding, is ubiquitous in the animal kingdom (Howard 1960; Clobert *et al.* 2001). This crucial life history trait has a strong impact on population genetic structure, influencing the maintenance and loss of genetic diversity in populations (Chesser 1991b; Sugg *et al.* 1996; Storz 1999). Hence,

Correspondence: Natasha Arora, Fax: +41 44 635 68 04; E-mail: n.arora@aim.uzh.ch

resolving a species' dispersal pattern as well as the mechanisms that drive these is of great interest.

Some of the evolutionary mechanisms invoked to explain the tendency for one sex to exhibit site fidelity or philopatry, i.e. the tendency to breed within or in close proximity to the natal range, include ecological benefits. For instance, philopatric individuals might benefit from familiarity with resources and avoid the risks associated with migration through unknown areas (Greenwood 1980; Lawson Handley & Perrin 2007). Philopatry results in kin structures that might also confer social benefits because of nepotistic interactions, providing inclusive fitness benefits that could augment or even drive philopatry (Perrin & Goudet 2001; Lawson Handley & Perrin 2007). The prediction for species with the mate-defence mating systems prevalent among mammals is that females, who benefit most from acquaintance with a given territory, should be philopatric, with males dispersing to avoid kin competition and inbreeding (Greenwood 1980; Dobson 1982; Pusey 1987; Wolff 1993).

The social organization of group-living mammals has drawn particularly intense interest. In these species, the salient social interactions have prompted many genetic studies to investigate whether kin structures among same-sex members underlie social behaviours such as tolerance, cooperation, learning and cultural variation (spotted hyenas; Van Horn *et al.* 2004; chimpanzees, Lukas *et al.* 2005; horses; Cameron *et al.* 2009; chacma baboons, King *et al.* 2011). Far fewer studies, however, have examined relatedness patterns in nongregarious species. Nevertheless, individuals of nongregarious species may have 'social networks' (Charles-Dominique 1978), engaging in associations with neighbours, so opportunities for cryptic kin selection to operate exist (Hatchwell 2010). Consequently, the exploration of kin structures in such species may lead to important new insights.

The few genetic investigations to date of nongregarious mammals have concentrated on carnivores (raccoons, Ratnayeke *et al.* 2002; cougars, Biek *et al.* 2006; bears, Zedrosser *et al.* 2007) and rodents (woodrats, McEachern *et al.* 2007), as well as a few lemur species among the primates (Kappeler *et al.* 2002; Eberle & Kappeler 2006; Radespiel *et al.* 2009). Such studies have proven invaluable, as illustrated by the examination of the solitarily foraging grey mouse lemur, a species in which females allo-nurse in diurnal sleeping groups. The usage of genetic markers enabled Eberle & Kappeler (2006) to establish that allo-nursing females comprised close maternal relatives, thus providing strong evidence for kin-based communal breeding. In other species without such opportunities for association, nepotistic behaviour could nonetheless still occur albeit in less obvious ways, for instance through reduced aggression and increased tolerance towards relatives that might make settlement in familiar areas easier (Perrin & Goudet 2001; Hatchwell 2010).

Among the most enigmatic nongregarious species are the Asian great apes, the orang-utans (genus: *Pongo*). Like most other great apes, orang-utans have a fission–fusion social system. But they stand out as a result of their especially low levels of sociality (van Schaik 1999) and possibly different social organization. In orang-utans, behavioural evidence points to female philopatry and male-biased dispersal (Galdikas 1985b; Mitani 1989; van Schaik & van Hooff 1996; Delgado & Van Schaik 2000), while in African great apes and humans, female dispersal is common (Eriksson *et al.* 2006; Wilkins & Marlowe 2006; Douadi *et al.* 2007; Langergraber *et al.* 2007; Guschanski *et al.* 2008). Such a dispersal pattern might affect associations among individuals, which despite occurring infrequently, do take place (van Schaik 1999; Delgado & Van Schaik 2000).

Yet the pattern of sex-biased dispersal in orang-utan populations is not clear. Broad-scale studies show tighter geographical clustering of mtDNA compared to Y-chromosome haplotypes across the highly differentiated orang-utan populations (Arora *et al.* 2010; Nietlisbach *et al.* accepted), suggesting historical male-mediated gene flow. Nevertheless, three previously published local scale studies of contemporary dispersal examining relatedness within populations did not confirm this pattern. These studies were based on conventional genetic methodology relying on the comparison of average pairwise relatedness (r) estimates of adult females and adult males obtained using biparentally inherited microsatellite markers. The expectation is that the more philopatric sex comprising related individuals should have higher r values than the dispersing sex comprising immigrants (Prugnolle & de Meeus 2002; Lawson Handley & Perrin 2007). The relatedness comparisons of the three studies were indicative of dispersal of both sexes (Utami *et al.* 2002), philopatry of both sexes (Goossens *et al.* 2006) or male-biased dispersal (Morrogh-Bernard *et al.* 2011). Nevertheless, the inclusion of rehabilitants in the first study, habitat fragmentation in the second study and the smaller sample size in the third study might have been responsible for these differences. The discrepant behavioural and genetic results render the social organization of orang-utans unresolved. It is also unclear whether contemporary dispersal patterns are at odds with historical patterns. Determining whether orang-utans have kin structures and how these are linked to dispersal is crucial step before investigating the possible evolutionary mechanisms underlying the movement of individuals and genes, population genetic structure, and social behaviour.

The aim of the present study was to gain an insight into the dispersal and relatedness patterns of orang-utans, based on the ongoing long-term study at Tuanan Orang-utan Research Area, Borneo, Indonesia. We capitalized on the largest set of genetically characterized sexually mature individuals ($n = 40$) from a natural population of orang-utans to test genetic predictions based on field observations of female philopatry and male-biased dispersal. We included only sexually mature individuals because they have potentially already settled within the natal area or dispersed to breed (Prugnolle & de Meeus 2002; Lawson Handley & Perrin 2007). By complementing spatial and behavioural information, as well as genetic data from the maternally inherited mitochondrial DNA (mtDNA) and a panel of 24 autosomal microsatellite markers, we tested the following predictions:

1. *MtDNA diversity patterns.* Diversity levels are expected to be higher for males if they are the dispersing sex, reflecting their more varied maternal ancestries.
2. *Pedigree relationships.* The number of closely related dyads, and especially maternally related dyads, as estimated from a parentage-based pedigree reconstruction, is expected to be higher among females compared to males.
3. *Average pairwise relatedness estimates.* The estimates are expected to be higher among females than males, as the latter should comprise immigrants.

In addition to disentangling the dispersal patterns in orang-utans, we discuss the effects of sampling regime, life history traits and spatial distribution of individuals on relatedness estimation, which is especially significant when studying nongroup-living animals.

Materials and methods

Study population

Sampling was conducted in the Tuanan Orang-utan Research Area (2°09′ South; 114°26′ East), Mawas Conservation Area, Central Kalimantan, Indonesia. This site is located within a peat swamp forest of approximately 750 ha, accessible through grid-based trails. The orang-utan density estimate for the area is 4.25–4.5 individuals per km^2 (van Schaik *et al.* 2005). Females at this site have home ranges estimated to be 325 ha (±125 ha) (Wartmann *et al.* 2010; van Noordwijk *et al.* 2012). Among males, two morphs are found: flanged males, which have fully developed irreversible secondary sexual characteristics, and unflanged males, which have not (Delgado & Van Schaik 2000; Utami *et al.* 2002). Home ranges of both flanged and unflanged males are far larger than those of females, also exceeding the size of the study site; their sizes are, therefore, unknown (Utami Atmoko *et al.* 2009; van Noordwijk *et al.* 2012).

Behavioural, spatial and genetic data collection

At this longitudinal study site, an intensive sampling regime from 2003 to 2009 targeted the collection of combined behavioural, spatial and genetic data for each individual, following the standard orang-utan protocol (http://www.aim.uzh.ch/orangutannetwork/FieldGuidelines.html). Trained observers conducted over 25 000 h of focal follows, normally nest-to-nest, to record behavioural and spatial information including space use, frequency of sightings, sex and age (Wich *et al.* 2004; van Noordwijk *et al.* 2012). The age of individuals born after 2003 was either known or estimated to the closest year; for individuals born before 2003, age was estimated based on known landmark ages in orang-utans (Wich *et al.* 2004).

Faecal samples were obtained during focal follows of individuals. Multiple samples were collected per individual throughout the study period and throughout the entire study area. We extracted DNA from the faecal samples with the QIAamp DNA Stool Mini Kit (Qiagen) and followed the manufacturer's protocol with a slight modification: elution was preceded by a 30-min incubation period. We genotyped individuals at up to 24 autosomal microsatellite markers and sequenced 450 bp of the hypervariable region I (HVRI) of the mtDNA using the same procedures as described in Arora *et al.* (2010).

For the genotyping, we minimized the genotyping errors associated with low quantity and quality of DNA obtained from noninvasively collected samples through the approach established by Morin *et al.* (2001). This method involves DNA quantification in each extract through real-time quantitative polymerase chain reaction (rtPCR), so as to determine the number of positive PCR replicates required to achieve a 99% certainty in a homozygous genotype. For a heterozygous genotype, the observation of each of the two alleles at least twice in independent PCRs is required. We initially used a panel of six autosomal microsatellite markers to genotype all samples obtained from potentially distinct individuals (Table S1, Supporting information). These markers were chosen because of their low-cumulative nonexclusion probabilities: 1.36×10^{-5} for unrelated individuals and 8.90×10^{-3} for full siblings, as determined by Cervus 3.0 (Kalinowski *et al.* 2007). Usage of these markers allowed us to distinguish unique individuals, providing a genetic method to link the behaviour of followed individuals to their genetic identity in a

longitudinal study. When repeated genotypes were obtained, we discarded all but one to have a data set of distinct individuals. These unique individuals were further genotyped at an additional 18 loci, resulting in a total of 24 autosomal microsatellite markers (Table S1), which were all in Hardy–Weinberg equilibrium, and showed no evidence of linkage disequilibrium or null alleles, as tested using Arlequin 3.11 (Excoffier *et al.* 2005), GenePop 4.0 (Rousset 2008) and ML-NullFreq (Kalinowski & Taper 2006), respectively. Details on the primers and PCR amplification conditions are described in the supporting information. For seven adult males, low autosomal DNA quality and quantity allowed only partial genotypes, restricted to the six markers used in the identity analyses. In total, multi-locus autosomal genotypes were obtained for 19 females and 29 males.

To obtain haplotype information, we sequenced 450 bp of the hypervariable region I (HVRI) of the mtDNA. Details on the primers, PCR amplification and raw data analyses are given in the Supporting information. MtDNA haplotypes were available for all genotyped individuals as well as one additional male with a unique mtDNA haplotype but no autosomal genotype.

Statistical analyses

We carried out the following analyses: (i) mtDNA diversity patterns, (ii) spatial distribution of females, whose ranging can be followed, (iii) parentage-based pedigree reconstruction and (iv) relatedness estimates. Unless specified otherwise, the analyses included only adult individuals who had potentially already settled within the natal area or dispersed to breed (Prugnolle & de Meeus 2002; Lawson Handley & Perrin 2007), the potential postdispersal (PPD) individuals. We considered individuals as PPD if they were sexually mature and/or regularly seen to range independently from the mother from the beginning of the study period (i.e. ranging at more than 50 m distance for at least several consecutive days). Individuals maturing during the study period were not included as PPD. These criteria resulted in a total of 40 PPD individuals ($n_{females}$ = 15; n_{males} = 25). The number of individuals included in each of the analyses detailed later is summarized in Table S3.

MtDNA diversity patterns and spatial distribution. Using the HVRI haplotypes, we conducted several analyses to assess patterns of mtDNA diversity and lineage relatedness. First, we compared levels of nucleotide and haplotype diversity for the PPD females and males using DNAsp v.5.0 (Librado & Rozas 2009). We tested for a significant difference in haplotype diversity between the sexes using a randomization test. For this, we randomly assigned all observed haplotypes to all males and females 1000 times and counted the number of instances in which the difference between male and female haplotype diversity exceeded the observed one. To show the mutational distances between the haplotypes found in the population as well as their frequencies according to sex, we generated a median-joining network using Network v4.6 and Network Publisher v1.2.0 (Bandelt *et al.* 1999; http://www.fluxus-engineering.com). Second, we assigned individuals to mtDNA lineages, defining these on the basis of haplotype sharing, irrespective of the biparental kinship of individuals.

For the PPD females, we also investigated the spatial distribution of mtDNA lineages using ArcGIS v.9.3.1 (ESRI 2008). To illustrate the areas within the study site where females with the same mtDNA haplotype, i.e. mtDNA lineages, were observed, we used the HRT plug-in for ArcGIS (Rodgers *et al.* 2007) to calculate 95% kernel probability plots, aggregating spatial data for all females with the same haplotype. Hence, incomplete ranging data for the females who also frequently moved outside of the study area did not affect the analyses. Spatial data were available for 13 PPD females (see Supporting information).

Parentage-based pedigree analyses. We examined the precise genetic relationships of female–female, male–male and female–male dyads through a combination of parentage and mtDNA analyses. First, we used the likelihood-based approach as implemented in Cervus 3.0 (Kalinowski *et al.* 2007) to carry out a parentage analysis for all PPD individuals for which data for 24 microsatellite markers were available ($n_{females}$ = 15; n_{males} = 17), as well as nine dependent offspring (see Supporting information).

Simulations were conducted to determine critical values of the log-likelihood score for a 95% confidence parentage assignment. The parameters for these simulations were 10,000 cycles and a minimum of 10 loci typed. The specified genotyping error rate of 0.112% was determined through the 'repeat-genotyping' and 'unintentionally re-sampled individuals' approaches described by Hoffman & Amos (2005). Only PPDs were incorporated as candidate mothers or fathers. The proportion of candidate parents was difficult to estimate from field data. Given the large influence this may have on the statistical significance of the results (Krützen *et al.* 2004a), several conservative values for this parameter (0.05, 0.08 and 0.10) were tested to check the robusticity of assignments. To examine the genetic relationships among all individuals including the seven additional males for which only a panel of six microsatellite markers was available, we repeated the parentage analyses with the same parameters, but with a specification of a minimum of five loci typed.

Following the parentage assignments, we inferred maternal and paternal sibling relationships by examining the shared mothers and shared fathers for each individual in the data set (n = 48). Such a parentage-based pedigree reconstruction allowed assessment of the number of maternal and paternal relatives at the site for each individual, incorporating parent-offspring and sibling relationships. These numbers represent only a minimum bound because the inference of genealogical relationships requires assignment to a parent and hence sampling of this parent within the study site, which may be limited by factors including emigration or death.

Relatedness analyses. We estimated average pairwise relatedness (r) coefficients for all PPD males and females in the data set. Two analyses were carried out. First, r was estimated for all same-sex individuals. Second, r was estimated for each set of same-sex individuals sharing their mtDNA haplotype.

To calculate r estimates, we used the triadic likelihood estimator (TrioML; Wang 2007). This estimator computes relatedness of a dyad in relation to a third reference individual in order to reduce errors stemming from identity-in-state rather than identity-by-descent. It further allows the specification of a genotyping error rate and is bounded between 0 and 1, a more legitimate range than that of other estimators. Moreover, an evaluation using empirical and simulated data for seven different estimators showed that the TrioML produced overall the most accurate estimates (Wang 2007). All PPD individuals were used as the reference population for the background allele frequency calculation. We compared the average relatedness between female dyads and male dyads and tested for significance through 1000 bootstrap re-samplings of the individuals from the observed data set and comparison of the differences in the observed and re-sampled data sets. To show deviations from the population mean, the r estimates were corrected by calibrating the population mean to zero.

In addition, and for comparative purposes, r estimates were also computed with three other estimators: (i) the coefficient of Queller & Goodnight (1989), which is frequently used in the literature and (ii) the coefficients of Wang (2002) and (iii) Lynch & Li (Lynch 1988; Li *et al.* 1993) chosen on the basis of their performance in an estimator evaluation conducted as detailed in the Materials and methods and Supporting information.

Results

Statistical analyses

MtDNA diversity patterns and spatial distribution

We investigated mtDNA diversity and haplotype-sharing patterns. In total, we found 10 different mtDNA haplotypes in Tuanan (Fig. 1; see Supporting information), with an overall, haplotype diversity h of 0.66 (SD ± 0.081) and nucleotide diversity π of 0.006 (SD ± 0.002). Two haplotypes were specific to females: haplotype B was found in four females (10% of individuals) and haplotype C in two females (5%). Another haplotype (A) was very common, found in 23 individuals, and shared by both males (35%) and females (22.5%). The other five haplotypes were all male-specific: haplotype D was present in three males (7.5%), haplotypes E and I in two males each (5%), and haplotypes F, G, H and J in one male each. Interestingly, two of the rare haplotypes unique to the males differed by at least nine mutational steps from the other haplotypes (Fig. 1). The mtDNA variation between the sexes led to a ten-fold higher mtDNA nucleotide diversity in males (π = 0.01 ± 0.002) compared to females (π = 0.001 ± 0.0003). The randomization procedure revealed a significantly higher haplotype diversity in PPD males compared to PPD females (Δ obs ($h_{m/f}$) = 0.102, P = 0.008). Both the presence of sex-specific haplotypes and the significantly higher haplotype diversity in males compared to females are consistent with the genetic predictions for female philopatry and male-biased dispersal.

We were also able to examine the spatial distribution of females, since their home ranges are smaller than those of males and than the study area. While females with haplotypes B and C have their home ranges mainly within the study site, the females with haplotype A range partly in the periphery. Nonetheless, the analyses show

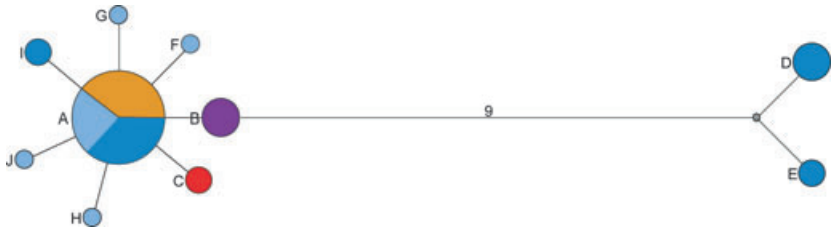


Fig. 1 MtDNA haplotypes in Tuanan. A median joining network of mtDNA haplotypes found in Tuanan is shown. Each different haplotype, shown as a circle, is coloured to represent the proportion of individuals sharing a haplotype: dark blue (flanged males), light blue (unflanged males), other colours (females). Number of mutations between haplotypes is one unless specified.

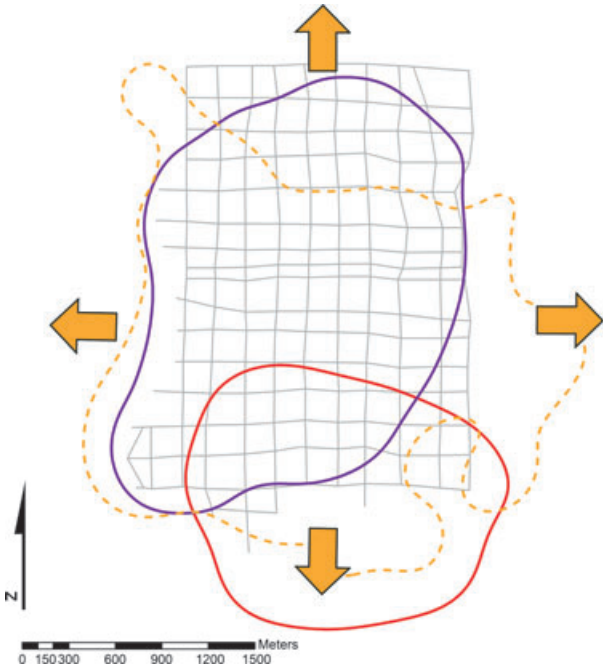


Fig. 2 Spatial distribution of mtDNA lineages in Tuanan. The grid represents the study site, with the combined ranges of females with the same haplotype represented by lines, colour-coded following Fig. 1. The dashed line corresponds to females that frequently moved out of study area (as highlighted by the arrows).

that, within the study site, there is extensive overlap of different mtDNA lineages, indicating that females with different haplotypes share space (Fig. 2).

Parentage-based pedigree analyses

Through the reconstruction of parentage-based pedigrees, we were able to examine the distribution of maternal and paternal relatives among females and males (Table 1). All maternal relationships were confirmed by the observed haplotype sharing. We found that 10 of 15 PPD females had a mother or a PPD daughter at the study site, while only 1 of 24 PPD males was assigned a mother, supporting a model of female philopatry and male-biased dispersal. Particularly the females ranging fully or largely within the study area, those with haplotypes B and C, formed clusters of related individuals. Our results indicate that cluster B comprises a mother and her three adult daughters, two of which in turn have adolescent female offspring. The two PPD females of cluster C were confirmed as a mother–daughter pair. Among the nine PPD females of cluster A, most of which range partly in the periphery of the study site, two mother–daughter pairs were found. The only PPD female with haplotype A and a home range mainly within the study area was not found to have PPD relatives in the area. Field observations indicate that this female had gradually moved from the disturbed habitat in which she had formerly ranged and was consistently chased away at every encounter with other PPD females. None of the males shared haplotypes with the well-known females from clusters B and C, indicating that this is not their natal area. No fathers or paternal relatives were assigned to any of the PPD females or males, indicating that the fathers of adult individuals are not likely to be in the study area.

Relatedness analyses

The average pairwise relatedness estimate *r* as computed with the TrioML estimator was significantly higher among females than males (P value <0.05; Fig. 3). This result was independent of the estimator used, as observed in the comparison across estimators (Fig. S1, Supporting information). We also estimated biparental relatedness for same-sex individuals from the same mtDNA lineage using the TrioML estimator (Fig. 3). The *r* estimates for females with the same mtDNA haplotype were higher than those obtained when all females were pooled together. Males sharing an mtDNA haplotype, by contrast, did not show higher biparental relatedness than all males, irrespective of haplotype. Among individuals with haplotype A, relatedness among females was also significantly higher than that among males.

Discussion

We integrated spatial, observational and genetic data to investigate the dispersal pattern in a nongregarious

Table 1 Maternal and paternal relatives of females and males at Tuanan

		With maternal relatives				With paternal relatives			
Sex	*N*	Mother (%)	Daughter/Son (%)	Sister (%)	Brother (%)	Father (%)	Daughter/Son (%)	Sister (%)	Brother (%)
Females	15	6 (40)	4 (27)	3 (20)	1 (6)	0 (0)	–	0 (0)	0 (0)
Males	24*	1 (4)	–	1 (4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

*For seven of the males, autosomal genotypes were available for the six loci used in the identity analyses, determined to be powerful for parentage assignments in assessments of marker informativeness (see Supporting Information).

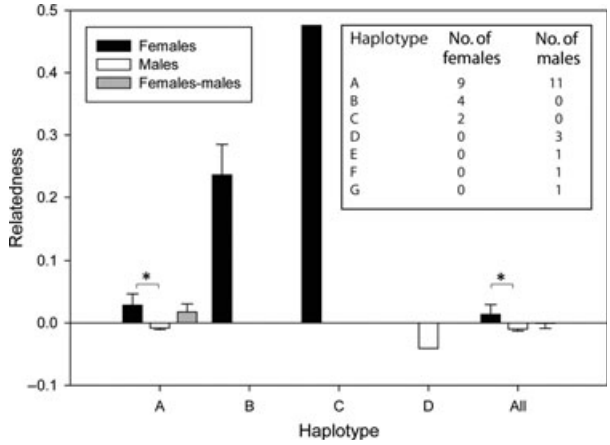


Fig. 3 Female and male biparental relatedness. Trio ML relatedness estimates corrected for population average, as well as variances (error bars) are shown for: all same-sex individuals, and same-sex individuals sharing an mtDNA haplotype. The statistically significant differences in relatedness (P-value < 0.05) are represented by asterisks. For each haplotype, the number of PPD females and PPD males for which complete autosomal genotypes were obtained (microsatellite markers) is detailed in the embedded table.

mammal for which previous genetic studies had produced mixed results. We tested relatedness patterns among individuals through three conventional genetic analyses: mtDNA diversity and haplotype-sharing analyses, the reconstruction of genealogical relationships through parentage analyses and average biparental relatedness. Our results revealed the presence of sex-specific haplotypes and significantly higher mtDNA diversity among males compared to females, underscoring the divergent maternal ancestries of the males. The average pairwise relatedness estimate was higher for females than for males. More importantly, while two-thirds of all females in our study had maternal relatives, with only one case this was the exception for males, indicating a pronounced pattern of female philopatry and male-biased dispersal.

Female philopatry and male-biased dispersal: evidence and comparisons

Among the females, we found three different mtDNA lineages containing clusters of close maternal relatives. For the females with haplotypes B and C, whose home ranges were mainly within the study site, we were able to fully disentangle maternal relationships. Although the relationships among females with haplotype A, most of which range peripherally, are less complete, we did detect two PPD mother–daughter pairs that ranged in the periphery of the study area.

While the female philopatric tendencies supported by our results are congruent with the dispersal patterns in some other solitarily foraging primates, some marked differences are apparent. Notably, there is extensive overlap in home ranges among these females, resulting in spatially stacked matrilineal clusters. These stacked matrilineal clusters contrast with the more spatially distributed maternal lineages in, for example, Coquerel's dwarf lemurs. For this species, Kappeler *et al.* (2002) showed that the sighting centres of females from the same mtDNA lineage are closer than those of females from different lineages. Moreover, within orang-utan clusters, we provided evidence that females are mainly first and second-degree relatives, comprising families of adult mothers and their adult daughters as well as their offspring, while the precise genealogical relationships of females in other nongregarious species are often not known or taken into account, although they may affect nepotistic interactions.

The males in this study, however, differed from the females in several ways. First, they had a far higher diversity of mtDNA haplotypes, most of which were sex-specific. The seven rare haplotypes pertaining exclusively to males highlight their different maternal ancestry compared to the females. This pattern of male-specific haplotypes mirrors the results of studies in the grey mouse lemur and Coquerel's dwarf lemur (Kappeler *et al.* 2002; Wimmer *et al.* 2002; Fredsted *et al.* 2004). Second, males rarely had first-degree relatives in the study area. It was especially revealing that males did not have any mothers or maternal sisters in the study area, except in the case of one young, probably predispersal, male. As none of the males shared maternal ancestry with the well-known centrally located females from clusters B and C, our results indicate that the study site is not a natal area for any of the males. In addition, data on the number of new distinct individuals identified each year indicate that new males keep coming into the study site, while the females are limited in number and well-known after a few years (Fig. S3, Supporting information).

Together, our results match the predictions for a model of female philopatry and male-biased dispersal, in line with previous studies of historical gene flow patterns (Arora *et al.* 2010; Nater *et al.* 2011) and behavioural observation at several orang-utan research sites (Galdikas 1985a; Mitani 1989; van Schaik & van Hooff 1996; Delgado & Van Schaik 2000). Our findings also agree with a recent broad-scale study comparing mitochondrial and Y-linked genetic markers, which provided evidence that orang-utan males move much further than females (Nietlisbach *et al.* in press).

Nevertheless, the patterns we found do not dismiss possible variation in the distances travelled by males, nor a potential range expansion. Some males shared the common haplotype A with the females ranging partly

outside the study area. Thus, it is possible that, unless haplotype A is extremely widespread in the population, these males have their maternal relatives not too far from the study area, suggesting that they have travelled short distances. As male ranges are large and surpass the size of the study site, it is not fully clear whether the males with haplotype A have home ranges that include their natal area, and if so, whether this feature is permanent or temporary, i.e. restricted to early stages of dispersal. Thus, there is a possibility that males with the common haplotype A have expanded their natal ranges, as occurs for instance with bottlenose dolphins (Krützen *et al.* 2004b).

In some cases, males shared their mtDNA haplotype with each other and thus could be maternally related, despite the negative r for males sharing a haplotype as compared to the population mean. Because a parentage-based pedigree reconstruction requires sampling the shared mother to make inferences on shared sibship, inferences on their genealogical relationships cannot be made. However, even if these males were maternally related, parallel male dispersal is unlikely given low male sociality (Delgado & Van Schaik 2000; Utami Atmoko *et al.* 2009). It is nonetheless possible for related males sharing maternal ancestry to converge at a site if the dispersal options are limited because of forest fragmentation and other ecological barriers. This is unlikely to hold for Tuanan, but may be an important consideration elsewhere.

Another interesting finding was that some of the rare sex-specific mtDNA haplotypes were found among unflanged males, who have not yet developed the irreversible secondary sexual characteristics found in the generally older flanged males (Delgado & Van Schaik 2000; Utami *et al.* 2002). Thus, in contrast to suggestions by Morrogh-Bernard *et al.* (2011), our findings indicate that male dispersal may occur when individuals are still young.

Factors affecting the power to disentangle dispersal patterns

Our investigation highlights the importance of several factors affecting the sensitivity of genetic approaches to measure dispersal, particularly for nongregarious species: sampling regime, life history traits and the spatial distribution of individuals.

First, we were better able to resolve the pedigree of females whose home ranges were fully or largely within the study site (cluster B and C), compared to that of females who only partially ranged within it (cluster A). This finding points to the critical importance of size of the sampling area relative to home range size, particularly in nongregarious species. While group-living species have cohesive distinct units of regularly interacting individuals that determine which individuals are sampled, the absence of such units in nongregarious species means that sampling is opportunistic, i.e. spatial rather than group-based criteria, resulting in potential discrepancies between behavioural and genetic results. Especially, the widely used average biparental relatedness estimates are subject to biases stemming from such opportunistic sampling. Species with relatively small home range sizes and small dispersal distances, relative to the sampling area, allow researchers to incorporate larger sample sizes. However, such a sampling regime might lead to the inclusion of unrelated members of the philopatric sex, resulting in lower r estimates than expected. One solution in this case is to measure genetic relatedness against spatial distance (Prugnolle & de Meeus 2002), as has also been performed for various group-living species (i.e. red deer, Nussey *et al.* 2005). To date, studies of a number of nongregarious small-distance travelling mammals show, in agreement with patterns of female philopatry, the expected decrease in female r estimates with increasing geographical distance, and little or no distance effect for males (Coquerel's dwarf lemurs; Kappeler *et al.* 2002; raccoons; Ratnayeke *et al.* 2002; Quail ridge woodrats; McEachern *et al.* 2007). Nevertheless, this approach is not always possible, especially for species with relatively large home ranges and large dispersal distances. Including individuals whose home ranges are not fully encompassed within a study area will reduce the genetic power to detect philopatry if these individuals have their relatives elsewhere.

Second, the slow life histories of some species such as orang-utans and other great apes lead to small sets of closely related individuals at a given time. Thus, in contrast to species with faster life histories, a given sampling area may contain lower numbers of related individuals among the philopatric sex, depending on home range size. It is to be expected then that r estimates decrease with increasing numbers of individuals included in an analysis, as observed in a study of chimpanzees (Lukas *et al.* 2005). Levels of relatedness will also vary depending on reproductive skew, with higher coancestry among the offspring sired by a male with high mating monopolization for instance (Chesser 1991a).

Third, we found stacked matrilineal clusters of females, whose home ranges overlapped. This spatio-genetic structure among females makes it difficult to assess relatedness, as there are both closely related dyads as well as unrelated dyads sharing the same area. This may have been a confounding factor in previous genetic studies of orang-utans, as estimates of average relatedness alone are poor measures of female philopatry. Such spatio-genetic structuring could also

explain cases in other species where, despite behavioural and genetic evidence for female philopatry, average relatedness for females is not higher than expected by chance, as in a study of cougars (Biek *et al.* 2006).

Nonetheless, there may still be some differences in the dispersal patterns across orang-utan populations as a result of intra-specific variation. Such variation would be indicative of facultative dispersal and a high degree of flexibility dependant on population density, local mate and resource competition, and in some cases kin cooperation. In the dusky-footed woodrat, for example, evidence for female kin structures was strongest at intermediate population densities, leading the authors to propose that 'high densities erode kin structures in response to local competition' (McEachern *et al.* 2007). In the grey mouse lemur, despite female philopatry, there is also evidence for the occasional dispersal of females. This was suggested by the spatial conglomeration of females with diverse haplotypes and no obvious female structuring, as well as the presence of multiple clusters of females that were not in spatial proximity but shared the same haplotype (Fredsted *et al.* 2004). In chimpanzees, despite the general pattern of extreme male philopatry and female-biased dispersal, recent research shows great variation in *r* across sites as well as in time (Mitani *et al.* 2002; Nishida *et al.* 2003; Lukas *et al.* 2005). Whereas at sites such as Mahale and Taï, almost all young females emigrate, at Gombe only 50% do and at Bossou none at all. The difference at the latter two sites has been attributed to their lower population sizes and greater isolation from other sites (Mitani *et al.* 2002; Nishida *et al.* 2003). In gorillas, males can either remain in the natal group or leave, and the fitness consequences of dispersal decisions for males at least have been shown to depend partly on demographic variables (Robbins & Robbins 2005). Another interesting possibility is that the recent availability of suitable unsettled habitat, as sometimes accompanies spatial expansions, could change the benefits and costs of dispersal, for instance, by increasing the fitness of dispersers.

Matrilineally related kin structures in orang-utans might confer social benefits to females. Despite the low levels of sociality displayed by orang-utans, associations do occur and have been shown to be more likely among related than unrelated females (van Noordwijk *et al.* 2012). Such associations provide opportunities for play among the offspring of closely related females (van Noordwijk *et al.* 2012). Taken together, these findings support suggestions by Singleton & van Schaik (2002) for the role of nepotistic tolerance in determining the nature of social interactions and opportunities to acquire new skills. Nepotistic tolerance might also make settlement in overlapping home ranges easier for relatives than nonrelatives. Given these results, kin selection may be an important evolutionary mechanism underpinning matrilineal kin structures not only in orang-utans but also in other nongregarious species where these structures remain underexplored, and which warrant detailed investigation.

Acknowledgements

We thank Jinliang Wang for his comments on the manuscript. We are also grateful to Pascal Marty and Claude Rosselet for their help with GIS work. This research would not have been possible without the help of the Indonesian Institute of Sciences and the Indonesian State Ministry for Research and Technology for granting permission to undertake this research. We also thank the A.H. Schultz foundation for additional funding. We are especially indebted to all assistants and students who helped collect and coordinate samples and behavioural data in the field. Samples were exported from Indonesia to Switzerland under permits 07279/IV/SATS-LN/2009, 00961/IV/SATS-LN/2007 from the Convention on International Trade in Endangered Species of Wild Fauna and Flora.

References

Arora N, Nater A, van Schaik CP *et al.* (2010) Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (Pongo pygmaeus). *Proceedings of the National Academy of Sciences, USA*, **107**, 21376–21381.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Biek R, Akamine N, Schwartz MK *et al.* (2006) Genetic consequences of sex-biased dispersal in a solitary carnivore: yellowstone cougars. *Biology Letters*, **2**, 312–315.

Cameron EZ, Setsaas TH, Linklater WL (2009) Social bonds between unrelated females increase reproductive success in feral horses. *Proceedings of the National Academy of Sciences*, **106**, 13850–13853.

Charles-Dominique P (1978) Solitary and gregarious prosimians:evolution of social structure in primates. In: *Recent Advances in Primatology* (eds Chivers DJ and Joysey KA), pp. 139–149. Academic, London.

Chesser RK (1991a) Gene diversity and female philopatry. *Genetics*, **127**, 437–447.

Chesser RK (1991b) Influence of gene flow and breeding tactics on gene diversity within populations. *Genetics*, **129**, 573–583.

Clobert J, Danchin E, Dhondt AA, Nichols JD (2001) *Dispersal*. Oxford University Press, New York.

Delgado RA, Van Schaik CP (2000) The behavioral ecology and conservation of the orangutan (Pongo pygmaeus): a tale of two islands. *Evolutionary Anthropology: Issues, News, and Reviews*, **9**, 201–218.

Dobson FS (1982) Competition for mates and predominant juvenile male dispersal in mammals. *Animal Behaviour*, **30**, 1183–1192.

Douadi MI, Gatti S, Levrero F *et al.* (2007) Sex-biased dispersal in western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology*, **16**, 2247–2259.

Eberle M, Kappeler PM (2006) Family insurance: kin selection and cooperative breeding in a solitary primate (*Microcebus murinus*). *Behavioral Ecology and Sociobiology*, **60**, 582–588.

Eriksson J, Siedel H, Lukas D *et al.* (2006) Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (*Pan paniscus*). *Molecular Ecology*, **15**, 939–949.

ESRI (2008) *ArcGis 9.3*. Environmental Systems Research Institute, Inc, Redlands, CA USA.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fredsted T, Pertoldi C, Olesen JM, Eberle M, Kappeler PM (2004) Microgeographic heterogeneity in spatial distribution and mtDNA variability of gray mouse lemurs (Microcebus murinus, Primates: Cheirogaleidae). *Behavioral Ecology and Sociobiology*, **56**, 393–403.

Galdikas BMF (1985a) Adult male sociality and reproductive tactics among orangutans at tanjung puting. *Folia Primatologica*, **45**, 9–24.

Galdikas BMF (1985b) Orangutan sociality at tanjung-puting. *American Journal of Primatology*, **9**, 101–119.

Goossens B, Setchell JM, James SS *et al.* (2006) Philopatry and reproductive success in Bornean orang-utans (*Pongo pygmaeus*). *Molecular Ecology*, **15**, 2577–2588.

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, **28**, 1140–1162.

Guschanski K, Caillaud D, Robbins MM, Vigilant L (2008) Females shape the genetic structure of a gorilla population. *Current Biology*, **18**, 1809–1814.

Hatchwell BJ (2010) Cryptic kin selection: kin structure in vertebrate populations and opportunities for kin-directed cooperation. *Ethology*, **116**, 203–216.

Hoffman JI, Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, **14**, 599–612.

Howard WE (1960) Innate and environmental dispersal of individual vertebrates. *American Midland Naturalist*, **63**, 152–161.

Kalinowski ST, Taper ML (2006) Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conservation Genetics*, **7**, 991–995.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099–1106.

Kappeler PM, Wimmer B, Zinner D, Tautz D (2002) The hidden matrilineal structure of a solitary lemur: implications for primate social evolution. *Proceedings. Biological Sciences*, **269**, 1755–1763.

King AJ, Clark FE, Cowlishaw G (2011) The dining etiquette of desert baboons: the roles of social bonds, kinship, and dominance in co-feeding networks. *American Journal of Primatology*, **7**, 768–774.

Krützen M, Barre LM, Connor RC, Mann J, Sherwin WB (2004a) 'O father: where art thou?'—Paternity assessment in an open fission-fusion society of wild bottlenose dolphins (Tursiops sp.) in Shark Bay, Western Australia. *Molecular Ecology*, **13**, 1975–1990.

Krützen M, Sherwin WB, Berggren P, Gales N (2004b) Population structure in an inshore cetacean revealed by microsatellite and mtDNA analysis: Bottlenose dolphins (Tursiops sp.) in Shark Bay, Western Australia. *Marine Mammal Science*, **20**, 28–47.

Langergraber KE, Siedel H, Mitani JC *et al.* (2007) The genetic signature of sex-biased migration in patrilocal chimpanzees and humans. *PLoS One*, **2**, e973.

Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology*, **16**, 1559–1578.

Li CC, Weeks DE, Chakravarti A (1993) Similarity of DNA fingerprints due to chance and relatedness. *Human Heredity*, **43**, 45–52.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lukas D, Reynolds V, Boesch C, Vigilant L (2005) To what extent does living in a group mean living with kin? *Molecular Ecology*, **14**, 2181–2196.

Lynch M (1988) Estimation of relatedness by DNA fingerprinting. *Molecular Biology and Evolution*, **5**, 584–599.

McEachern M, Eadie J, Van Vuren D (2007) Local genetic structure and relatedness in a solitary mammal, *Neotoma fuscipes*. *Behavioral Ecology and Sociobiology*, **61**, 1459–1469.

Mitani JC (1989) Orangutan activity budgets—monthly variations and the effects of body size, parturition, and sociality. *American Journal of Primatology*, **18**, 87–100.

Mitani JC, Watts DP, Muller MN (2002) Recent developments in the study of wild chimpanzee behavior. *Evolutionary Anthropology*, **11**, 9–25.

Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (Pan troglodytes verus). *Molecular Ecology*, **10**, 1835–1844.

Morrogh-Bernard HC, Morf NV, Chivers DJ, Krützen M (2011) Dispersal patterns of Orang-utans (Pongo spp.) in a Bornean Peat-swamp forest. *International Journal of Primatology*, **32**, 362–376.

Nater A, Nietlisbach P, Arora N *et al.* (2011) Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant orangutans (genus: Pongo). *Molecular Biology and Evolution*, **28**, 2275–2288.

Nietlisbach P, Arora N, Nater A *et al.* (in press) Heavily male-biased long distance dispersal of orang-utans (genus: *Pongo*), as revealed by Y chromosomal and mitochondrial genetic markers. *Molecular Ecology*.

Nishida T, Corp N, Hamai M *et al.* (2003) Demography, female life history, and reproductive profiles among the chimpanzees of Mahale. *American Journal of Primatology*, **59**, 99–121.

van Noordwijk MA, Arora N, Willems EP *et al.* (2012) Female philopatry and its social benefits among Bornean orangutans. *Behavioural Ecology and Sociobiology*. DOI: 10.1007/s00265-012-1330-7.

Nussey DH, Coltman DW, Coulson T *et al.* (2005) Rapidly declining fine-scale spatial genetic structure in female red deer. *Molecular Ecology*, **14**, 3395–3405.

Perrin N, Goudet J (2001) Inbreeding, kinship, and the evolution of natal dispersal. In: *Dispersal* (eds Clobert J, Danchin E, Dhondt AA and Nichols JD), pp. 123–142. Oxford University Press, New York.

Prugnolle F, de Meeus T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity*, **88**, 161–165.
Pusey AE (1987) Sex-biased dispersal and inbreeding avoidance in birds and mammals. *Trends in Ecology & Evolution*, **2**, 295–299.
Queller DC, Goodnight KF (1989) Estimating relatedness using genetic-markers. *Evolution*, **43**, 258–275.
Radespiel U, Jurić M, Zimmerman E (2009) Sociogenetic structures, dispersal and the risk of inbreeding in a small nocturnal lemur, the goldenbrown mouse lemur (*Microcebus ravelobensis*). *Behaviour*, **146**, 607–628.
Ratnayeke S, Tuskan GA, Pelton MR (2002) Genetic relatedness and female spatial organization in a solitary carnivore, the raccoon, Procyon lotor. *Molecular Ecology*, **11**, 1115–1124.
Robbins AM, Robbins MM (2005) Fitness consequences of dispersal decisions for male mountain gorillas (*Gorilla beringei beringei*). *Behavioral Ecology and Sociobiology*, **58**, 295–309.
Rodgers AR, Carr AP, Beyer HL, Smith L, Kie JG (2007) *HRT: Home Range Tools for ArcGIS*. Ontario Ministry of Natural Resources, Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario, Canada., Ontario, Canada.
Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
van Schaik C (1999) The socioecology of fission-fusion sociality in Orangutans. *Primates*, **40**, 69–86.
van Schaik CP, van Hooff JARAM (1996) Toward an understanding of the orangutan's social system. In: *Great Ape Societies* (eds McGrew WC, Marchant LF and Nishida T), pp. 3–15. Cambridge University Press, Cambridge.
van Schaik CP, Wich SA, Utami SS, Odom K (2005) A simple alternative to line transects of nests for estimating orangutan densities. *Primates*, **46**, 249–254.
Singleton I, van Schaik CP (2002) The social organisation of a population of Sumatran orang-utans. *Folia Primatologica*, **73**, 1–20.
Storz JF (1999) Genetic consequences of mammalian social structure. *Journal of Mammalogy*, **80**, 553–569.
Sugg DW, Chesser RK, Dobson FS, Hoogland JL (1996) Population genetics meets behavioral ecology. *Trends in Ecology & Evolution*, **11**, 338–342.
Utami Atmoko S, Singleton I, van Noordwijk M, van Schaik C, Setia T (2009) Male-male relationships in orang-utans. In: *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (eds Wich S, Utami Atmoko S, Setia T and van Schaik C)., pp. 225–233. Oxford University Press, New York.
Utami SS, Goossens B, Bruford MW, de Ruiter JR, van Hooff JARAM (2002) Male bimaturism and reproductive success in Sumatran orang-utans. *Behavioral Ecology*, **13**, 643–652.
Van Horn RC, Engh AL, Scribner KT, Funk SM, Holekamp KE (2004) Behavioural structuring of relatedness in the spotted hyena (*Crocuta crocuta*) suggests direct fitness benefits of clan-level cooperation. *Molecular Ecology*, **13**, 449–458.
Wang J (2002) An estimator for pairwise relatedness using molecular markers. *Genetics*, **160**, 1203–1215.
Wang J (2007) Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetical Research*, **89**, 135–153.
Wartmann FM, Purves RS, van Schaik CP (2010) Modelling ranging behaviour of female orang-utans: a case study in Tuanan, Central Kalimantan, Indonesia. *Primates*, **51**, 119–130.
Wich SA, Utami-Atmoko SS, Setia TM *et al.* (2004) Life history of wild Sumatran orangutans (*Pongo abelii*). *Journal of Human Evolution*, **47**, 385–398.
Wilkins JF, Marlowe FW (2006) Sex-biased migration in humans: what should we expect from genetic data? *BioEssays*, **28**, 290–300.
Wimmer B, Tautz D, Kappeler PM (2002) The genetic population structure of the gray mouse lemur (*Microcebus murinus*), a basal primate from Madagascar. *Behavioral Ecology and Sociobiology*, **52**, 166–175.
Wolff JO (1993) What is the role of adults in mammalian juvenile dispersal. *Oikos*, **68**, 173–176.
Zedrosser A, Stoen OG, Saebo S, Swenson JE (2007) Should I stay or should I go? Natal dispersal in the brown bear. *Animal Behaviour*, **74**, 369–376.

N.A., C.A., A.N., M.G., P.N., and M.K. apply genetic methods to study evolutionary, population genetic and ecological questions in wild animal populations. M.K. and E.P.W. investigate the evolution of culture and cooperation using genetic and spatial data. S.S.U.A., J.P., and D.P.W. study Indonesian primates. L.D., Mv.N. and Cv.S. are interested in the social evolution and cognition of primates.

Data accessibility

The mtDNA HVRI sequences have been deposited in EMBL under accession numbers FR717918–FR717919 and FR717921–FR717924. The genotypes and haplotypes for each individual, as well as the spatial distribution data for the females, are provided in the Supporting information.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Supporting Information Methods and Results.

Appendix S2 Supporting Information Tables and Figures.

Appendix S3 Supporting Information Genotypes and Haplotypes.

Appendix S4 Supporting Information Haplotypes.

Appendix S5 Supporting Information nexus file Tuanan New Haplotypes.

Appendix S6 Supporting Information Ranging Data.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 3173–3186 doi: 10.1111/j.1365-294X.2012.05539.x

Heavily male-biased long-distance dispersal of orang-utans (genus: *Pongo*), as revealed by Y-chromosomal and mitochondrial genetic markers

PIRMIN NIETLISBACH*†, NATASHA ARORA*, ALEXANDER NATER*, BENOIT GOOSSENS‡§, CAREL P. VAN SCHAIK* and MICHAEL KRÜTZEN*

*Anthropological Institute and Museum, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland, †Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland, ‡Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Biomedical Sciences Building, Museum Avenue, Cardiff CF10 3AX, UK, §Danau Girang Field Centre, c/o Sabah Wildlife Department, Wisma Muis, 88100 Kota Kinabalu, Sabah, Malaysia

Abstract

Mating systems are thought to be an important determinant of dispersal strategies in most animals, including the great apes. As the most basal taxon of all great apes, orang-utans can provide information about the evolution of mating systems and their consequences for population structure in this Family. To assess the sex-specific population structure in orang-utans, we used a combination of paternally transmitted Y-chromosomal genetic markers and maternally transmitted mitochondrial DNA sequences. Markers transmitted through the more philopatric sex are expected to show stronger differentiation among populations than the ones transmitted through the dispersing sex. We studied these patterns using 70 genetic samples from wild orang-utans from seven Bornean and two Sumatran populations. We found pronounced population structure in haplotype networks of mitochondrial sequence data, but much less so for male-specific markers. Similarly, mitochondrial genetic differentiation was twice as high among populations compared to Y-chromosomal variation. We also found that genetic distance increased faster with geographic distance for mitochondrial than for Y-linked markers, leading to estimates of male dispersal distances that are several-fold higher than those of females. These findings provide evidence for strong male-biased dispersal in orang-utans. The transition to predominantly female-biased dispersal in the great ape lineage appears to be correlated with life in multimale groups and may reflect the associated fitness benefits of reliable male coalitions with relatives or known partners, a feature that is absent in orang-utans.

Keywords: conservation, great apes, isolation by distance, male-specific markers, orang-utan, sex-biased dispersal

Received 26 October 2011; revision received 6 February 2012; accepted 10 February 2012

Introduction

Dispersal affects the population dynamics and genetics of species (Prugnolle & de Meeus 2002). On an individual level, dispersal may be costly, due to increased mortality because of unfavourable habitats, unfamiliarity with the new area, and loss of known or related cooperation partners (reviewed by Lawson Handley & Perrin 2007). Possible benefits include the acquisition of more resources in a new area, prevention of kin competition over mates and resources, as well as avoidance of inbreeding, which are usually achieved if there is a sex difference in dispersal. Most mammals show male-biased dispersal, which is usually explained by their predominantly polygynous mating systems (Greenwood 1980; Lawson Handley & Perrin 2007).

Correspondence: Michael Krützen, Fax: +41 44 635 61 12; E-mail: michael.kruetzen@aim.uzh.ch

Among great apes, different dispersal patterns have been described. Chimpanzees (*Pan troglodytes*; *e.g.* Nishida *et al.* 2003) and bonobos (*Pan paniscus*; *e.g.* Hohmann 2001) show strongly female-biased dispersal. In western lowland gorillas (*Gorilla gorilla gorilla*), both sexes may disperse (*e.g.* Harcourt *et al.* 1976), with dispersing males moving further than females (Douadi *et al.* 2007), but males may be more likely to stay in their group if they can inherit breeding status from their father (Watts 2000). There is currently no consensus as to the extent of sex-biased dispersal in humans (*Homo sapiens*). On a local scale, sex bias in dispersal patterns may be related to subsistence (Segurel *et al.* 2008), but differing sex biases can occur among otherwise similar villages (Oota *et al.* 2001). On a global scale, behavioural studies suggest predominantly female-biased dispersal in food-producing societies (reviewed in Burton *et al.* 1996) and slightly female-biased dispersal in hunter-gatherers (Hill *et al.* 2011), but genetic studies have yielded conflicting findings (Seielstad *et al.* 1998; Stoneking 1998; Wilder *et al.* 2004). Given this, it is of great interest to examine the dispersal patterns of the genus *Pongo*, which is basal to the great apes (Raaum *et al.* 2005) and comprises both currently recognized species of orang-utans (*P. pygmaeus* on Borneo, *P. abelii* on Sumatra). Male-biased dispersal in orang-utans would render female-biased dispersal a derived trait in chimpanzees, bonobos and humans.

Behavioural observations of orang-utans at several long-term study sites show a larger number of transient males than females and maturing females settling near their mothers (van Schaik & van Hooff 1996; Delgado & van Schaik 2000; van Noordwijk *et al.* 2012). Single males have also been observed roving far away from known orang-utan populations (Rijksen 1978). Although all these findings suggest male-biased dispersal in orang-utans, genetic studies to date have produced ambiguous results. Goossens *et al.* (2006b) found high mean relatedness levels within both adult sex classes at a Bornean site, but Utami and colleagues (2002) found the opposite pattern for a Sumatran population. It has been suggested that the results of these studies may be influenced by direct dispersal barriers due to anthropogenic pressure, in particular extensive logging and agriculture (Goossens *et al.* 2006b), as well as the inclusion of rehabilitated orang-utans that came from elsewhere (Utami *et al.* 2002), leading to artificially lowered relatedness values. Results from another study site recovering from logging pressure, but without significant dispersal barriers (Morrogh-Bernard *et al.* 2011), were consistent with male-biased dispersal, as average relatedness was significantly higher among six females compared to ten males, suggesting male-biased dispersal in at least this population. Two recent studies at another Bornean site, focussing on the benefits of female philopatry (van Noordwijk *et al.* 2012) and establishing relatedness patterns through a combination of spatial, genetic and behavioural patterns (Arora 2011), also corroborate that at least at this site, females are philopatric, whereas males disperse.

Even though behavioural and autosomal genetic data are useful to study sex-biased dispersal, the conclusions that can be drawn are limited in space and time. Behavioural sampling can determine whether one sex is more likely to leave the natal area than the other, but in the absence of information on actual dispersal distances, this information cannot provide information about its effects on the genetic population structure. Autosomal markers are biparentally inherited, leading to a rapid breakdown of sex-specific dispersal signals in a matter of one generation. Thus, autosomal markers only allow the investigation of instantaneous sex-specific dispersal patterns and may fail to resolve even highly biased dispersal patterns (Lukas *et al.* 2005). Additionally, the spatial limitations of field studies make it impossible to capture the overall historical pattern at the species-wide level.

Because of these limitations, there is still a considerable gap in our understanding of sex-biased dispersal and its long-term genetic consequences in the genus *Pongo*. We aimed to address these questions by applying both paternally and maternally transmitted genetic markers, in conjunction with extensive sampling in wild populations throughout the entire range of the genus, with a particular focus on Bornean orang-utans. The marker system transmitted strictly through the dispersing sex is expected to show less geographic structure and less differentiation between study sites (Oota *et al.* 2001). The differential effect of geographic distance on genetic distance measured with each marker type can inform us about differences in effective dispersal distance between sexes. Although contrasting maternally inherited mitochondrial DNA variation and paternally inherited Y-chromosomal variation is a promising approach to investigate sex-biased dispersal in mammals, this method has rarely been chosen due to the lack of polymorphic Y-chromosomal markers for most mammals (Luo *et al.* 2007; Greminger *et al.* 2010). Here, we capitalized on the recent development of species-specific Y-chromosomal markers (Nietlisbach *et al.* 2010) to address these questions in orang-utans.

Methods

Sampling strategy

Faecal samples were collected noninvasively from wild, mature orang-utan males at seven study sites on Borneo

and two on Sumatra (Fig. 1 and Table 1), following the sampling protocol provided at http://www.aim.uzh.ch/orangutannetwork/GeneticSamplingProtocol.html. Individual males were visually identified in the field or distinguished using autosomal microsatellites (Arora *et al.* 2010). We sampled maternally inherited mitochondrial DNA and paternally transmitted Y chromosomes from the same males, in order to include markers that were brought together in the same male by reproduction of its parents. The advantage of such an approach is that differences between marker systems due to potentially sex-biased (but undetectable) inclusion of outliers could cause interpretative problems in cases where mitochondrial and Y-chromosomal data were derived from different individuals (*e.g.* the female or male part of the sample could by chance include more animals that dispersed exceptionally far). On the other hand, our sampling strategy should slightly overestimate the dispersal signal for mitochondrial DNA, as mitochondrial DNA is sampled after dispersal of the male. Thus, for male-biased dispersal scenarios, the applied sampling strategy is conservative.

DNA extraction

Faecal samples were either stored in RNAlater Solution (Applied Biosystems) or 99.8% EtOH, or collected in 99.8% EtOH and later dried using silica gel (Nsubuga *et al.* 2004). DNA was extracted using the QIAamp DNA Stool Mini Kit (Qiagen) following the manufacturer's instructions for human DNA extraction from stool samples with the following modifications. After adding 1.6 mL buffer ASL and thorough vortexing, samples were left at room temperature for 20 min. The centrifugation step to pellet stool particles and inhibitors bound to the InhibitEX matrix lasted 8 minutes, and the final elution was conducted with 100 μL buffer AE and a centrifugation time of 5 minutes. For some samples, all the steps after the InhibitEX treatment were automatically performed by the QIAcube robotic workstation (Qiagen).

Genetic marker systems

We assessed the sex-biased dispersal by comparing paternally and maternally transmitted marker systems. As male-specific markers, we employed 11 human-derived Y-linked microsatellite loci (Erler *et al.* 2004; Kayser *et al.* 2004; Nietlisbach *et al.* 2010), six single-nucleotide Y-polymorphisms (SNPs) and one insertion–deletion Y-polymorphism (Nietlisbach *et al.* 2010). All 18 loci were amplified in two multiplex polymerase chain reactions (Nietlisbach *et al.* 2010). For each individual, we combined all 18 markers into a single Y-chromosomal haplotype. As a female-specific marker, we used 323 base pairs of the hypervariable region I (HVR-I) of the mitochondrial control region. Sequencing of HVR-I was performed as described elsewhere (Arora *et al.* 2010).

Genetic diversity

Haplotypic diversity for Y-chromosomal microsatellite and mitochondrial sequence data for nine study sites (Table 1) was calculated in Arlequin v3.5 (Excoffier & Lischer 2010) according to Nei's (1987) formula, which corrects for small sample size. To test whether mean haplotypic diversity per study site differed between mtDNA and Y-chromosomal data on Borneo, we performed a Wilcoxon rank sum test in R v2.12.1 (R Development Core Team 2010). To illustrate genetic diversity within study sites, we calculated the mean number of pairwise differences among haplotypes (π) for Y-linked and mitochondrial data, as well as the mean number of nucleotide differences (π_n) for mitochondrial sequences in Arlequin v3.5. Mean pairwise stepwise differences (MPSD) among Y-haplotypes, analogous to π, using the number of mutations under the stepwise mutation model as a proxy for haplotype differences, were calculated in R v2.12.1. These calculations were made on a marker-by-marker basis, therefore permitting the inclusion of individuals with partly missing data. Standard deviations of MPSDs were calculated using Tajima's (1993) sampling and stochastic variance in R v2.12.1.

Haplotype networks

To illustrate the similarity of haplotypes and visually investigate how the grouping of haplotypes may be related to the geographic proximity of study sites, we constructed median-joining networks (Bandelt *et al.* 1999) using Network v4.5.1.0 and Network Publisher v1.1.0.7 (http://www.fluxus-engineering.com/sharenet.htm) for both marker systems. The network based on Y-chromosomal data used all SNPs, indel and microsatellite data, while the mtDNA network was constructed using all 323 bases of the HVR-I. Only individuals for which we had complete haplotypes available were included (for sample sizes, see Table 1). One potential problem with haplotype network construction is that both the weighing scheme and the network parameter epsilon may have an undue influence on the way that nodes are connected and median vectors kept (Bandelt *et al.* 1999). We weighted Y-chromosomal loci 100 times their gene identity (Nei 1987) over all orang-utan samples for which we had complete haplotypes available, because highly variable loci should be weighted less

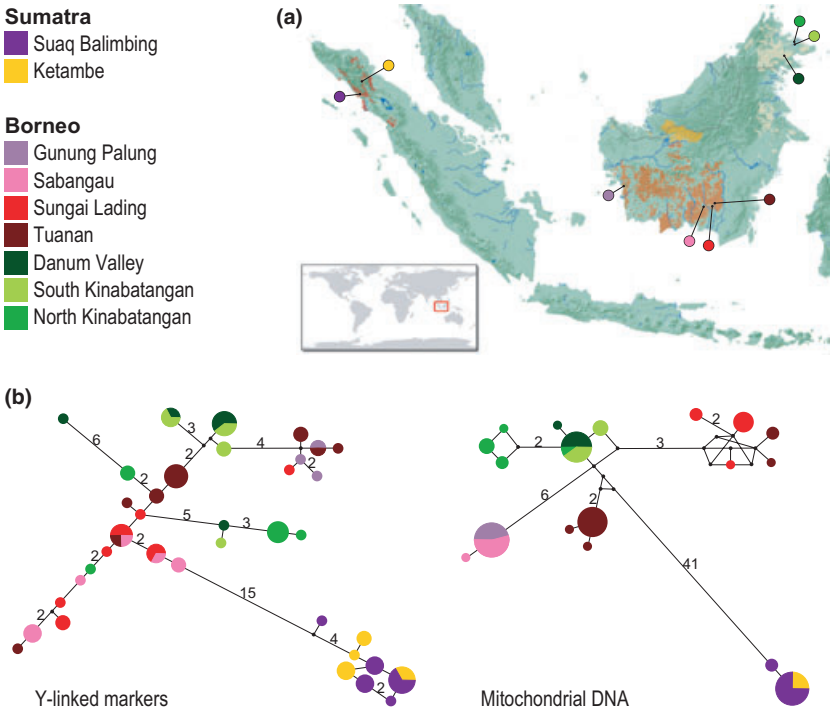


Fig. 1 Map of the study sites on Borneo and Sumatra and median-joining networks for Y-linked and mitochondrial markers. (a) The coloured areas on the map show the known distribution of the Sumatran orang-utan (red = *Pongo abelii*) and the three described subspecies of Bornean orang-utans according to Wich *et al.* 2008: orange = *Pongo pygmaeus wurmbii*; yellow = *P. p. pygmaeus*; grey = *P. p. morio*). (b) Median-joining networks for all Y-chromosomal markers and mitochondrial HVR-I sequences. Pies represent haplotypes and are coloured according to geographic origin and sized relative to the number of individuals carrying the respective haplotype. Numbers above lines represent the number of mutations occurring between two nodes, if higher than 1.

than more slowly evolving loci (Bandelt *et al.* 1999). Different weighting schemes were compared, but they did not significantly influence the resulting networks (data not shown). To assess the robustness, we qualitatively compared additional networks using only subsets of our Y-chromosomal markers (*e.g.* using mainly highly or less variable microsatellites and/or SNPs). We did not observe any character conflicts (data not shown). Similarly, networks obtained with high and low epsilon values (*i.e.* a tolerance up to which mutational distances among haplotypes are considered equal during the search for median vectors; Bandelt *et al.* 1999) were compared, but as high epsilon values and thus more reticulations did not show obvious character conflicts (data not shown), an epsilon value of zero was used for the presented networks.

Differentiation measures

Because of the small sample size for Sumatra, we restricted population genetic analyses to Borneo. We calculated R_{ST} for Y-microsatellites and φ_{ST} for mtDNA, under the assumption of a stepwise mutation and infinite allele model, respectively. Significance was assessed by generating a null distribution from 10 000 permutations, as implemented in Arlequin v3.5. Average differentiation among populations should be highest for the marker system transmitted through the more philopatric sex and was assessed by calculating several differentiation statistics. Previous work showed that for haplotypic data, dependence of differentiation measures on marker mutation rates is negligible when the applied statistic is based on the appropriate mutation model (Kronholm *et al.* 2010). Therefore, we calculated R_{ST} for the Y-linked microsatellites, based on the squared number of mutational steps among pairs of haplotypes and thus assuming a stepwise mutation model (Slatkin 1995). As differentiation measure for mitochondrial sequence data, we employed standard φ_{ST} statistics based on average pairwise differences between sequences (Excoffier *et al.* 1992). AMOVA-based calculations of both measures were made in Arlequin v3.5.

Because microsatellite markers often do not follow the stepwise mutation model (*e.g.* Masters *et al.* 2010), we also calculated a standardized differentiation measure (φ'_{ST}) by dividing observed φ_{ST} estimates by values that were maximized for the given within-population variation for both mitochondrial and Y-chromosomal data

Table 1 Diversity of Y-chromosomal microsatellites and mitochondrial HVR-I sequences for orang-utan males

		Y-linked microsatellites				Mitochondrial DNA (HVR-I)				*N* networks	
Study site	N	#ht	hd ± SD	π (MPD) ± SD	MPSD ± SD	#ht	hd ± SD	π (MPD) ± SD	π_n ± SD	Y	mtDNA
Borneo											
Danum Valley	6	3	0.600 ± 0.215	2.467 ± 1.544	5.867 ± 3.265	2	0.333 ± 0.215	0.333 ± 0.380	0.00103 ± 0.00136	6	6
Gunung Palung	6	5	0.933 ± 0.122	1.400 ± 0.990	1.933 ± 1.269	1	0	0	0	3	6
North Kinabatangan	7	3	0.667 ± 0.160	2.762 ± 1.658	6.581 ± 3.540	4	0.810 ± 0.130	1.619 ± 1.083	0.00501 ± 0.00384	8	7
Sabangau	8	5	0.857 ± 0.108	2.357 ± 1.433	4.000 ± 2.234	2	0.250 ± 0.180	0.250 ± 0.311	0.00077 ± 0.00110	8	8
South Kinabatangan	7	4	0.857 ± 0.102	2.857 ± 1.706	5.238 ± 2.881	2	0.571 ± 0.120	0.571 ± 0.521	0.00177 ± 0.00185	7	7
Sungai Lading	8	6	0.929 ± 0.084	2.857 ± 1.678	4.571 ± 2.511	3	0.607 ± 0.164	1.893 ± 1.203	0.00586 ± 0.00424	10	8
Tuanan	14	9	0.901 ± 0.062	2.560 ± 1.461	3.549 ± 1.919	5	0.593 ± 0.144	4.253 ± 2.243	0.01317 ± 0.00780	14	15
Mean Borneo	8.0 ± 2.8	5.0 ± 2.1	0.821 ± 0.133	2.466 ± 0.508	4.534 ± 1.554	2.7 ± 1.4	0.452 ± 0.272	1.274 ± 1.496	0.00394 ± 0.00463	8.0 ± 3.4	8.1 ± 3.1
Total Borneo	56	31	0.971 ± 0.009	3.319 ± 1.731	6.254 ± 3.014	16	0.884 ± 0.000	7.022 ± 3.348	0.02174 ± 0.01150	56	57
Sumatra											
Ketambe	3	2	0.667 ± 0.314	0.667 ± 0.667	1.333 ± 1.098	1	0	0	0	8	3
Suaq Balimbing	11	4	0.782 ± 0.075	0.727 ± 0.585	2.285 ± 1.355	2	0.327 ± 0.153	0.327 ± 0.355	0.00101 ± 0.00124	12	11
Mean Sumatra	7.0 ± 5.7	3.0 ± 1.4	0.725 ± 0.081	0.697 ± 0.042	1.809 ± 0.673	1.5 ± 0.7	0.164 ± 0.231	0.164 ± 0.231	0.00051 ± 0.00071	10.0 ± 2.8	7.0 ± 5.7
Total Sumatra	14	5	0.769 ± 0.083	0.692 ± 0.557	2.022 ± 1.209	2	0.264 ± 0.136	0.264 ± 0.308	0.00082 ± 0.00107	20	14

N = number of samples; the same adult male individuals were used for Y-chromosomal and mitochondrial data for all analyses except haplotype networks; sample sizes for haplotype networks are indicated in the last two columns; #ht = number of distinct haplotypes per study site; hd ± SD = haplotypic diversity with standard deviation; π (MPD) ± SD = mean number of pairwise differences; π (MPSD) ± SD = mean number of pairwise stepwise differences (mean number of mutational differences under the stepwise mutation model); π_n ± SD = nucleotide diversity (mean number of pairwise differences among nucleotides). Means ± SD over all study sites on each island are given, as well as total variation when treating each island as a single population.

(Hedrick 2005; Meirmans 2006; Langergraber *et al.* 2007b). We calculated 95% confidence intervals using 10,000 bootstrap replications over loci for the original and recoded (*i.e.* used only for maximization) data sets. The lower confidence limit of φ'_{ST} was calculated by dividing the lower limit obtained from the original data set by the upper value obtained from the recoded data set. Dividing the upper confidence limit obtained from the original data set by the lower limit obtained from the recoded data set yielded the upper confidence limit of φ'_{ST} (Langergraber *et al.* 2007b).

Comparing genetic differentiation between different marker systems with different substitution rates and levels of genetic variation may be problematic. F_{ST} statistics (Weir & Cockerham 1984) are influenced by the diversity of the markers used, and it has been shown that the theoretically possible value of full differentiation (= 1) is not reached even in cases of no allele sharing between populations, as long as there is variation within populations (Hedrick 2005; Jost 2008, 2009; Ryman & Leimar 2009). The measure D_{EST} had been proposed to replace F_{ST} whenever differentiation is the quantity of interest (Jost 2008, 2009). D_{EST} may be suitable for single-locus comparisons or multiple unlinked loci. However, D_{EST} is not useful for haplotypic data when most populations do not share haplotypes, because of full differentiation for most pairwise comparisons, irrespective of the similarity of haplotypes.

Genetic differentiation measured by φ_{ST} takes differences between haplotypes into account, but it also needs standardization in order to compare differentiation among different marker systems with different substitution rates (Hedrick 2005; Meirmans 2006). This can be achieved by creating a maximally divergent pseudo-data set based on the original patterns and dividing the original values by these maximal values, as described above (Hedrick 2005; Meirmans 2006). However, for the creation of the pseudo-data set, only variable sites are to be included, as invariable sites in the original data set will be transformed into sites still invariable within each population but variable between populations (Meirmans 2006; Langergraber *et al.* 2007b). Therefore, inclusion of such sites increases the values of the maximally diverse data set, making the standardization less efficient. To counter this problem, one could exclude invariable sites from the data set, but this does not solve the problem for multiple (three or more) populations, as there will usually be markers that are invariable between two populations but variable within the whole sample. Haplotypic marker systems with loci having few alleles or low variation are much more strongly affected by this bias. The reason is that the proportion of loci that is polymorphic within the whole data set but not within pairs of populations is much higher for such markers (such as DNA sequences or microsatellites with low variation) than for more variable multiallelic loci (such as highly variable microsatellites). Thus, for comparisons of different haplotypic marker systems involving multiple populations, we still lack an unbiased differentiation measure. The results of this study, however, were qualitatively identical for all mentioned measures of population structure and also quantitatively similar.

To investigate how genetic variation is distributed among geographically clustered groups of study sites, we conducted an AMOVA with a regional grouping (Table 3) of study sites as an additional hierarchical level. Arlequin v3.5 was used to calculate the AMOVA for Y-chromosomal microsatellites and mitochondrial sequence data. The analysis for mitochondrial haplotypes was based on a distance matrix of the number of different sites between pairs of samples. For Y-linked microsatellites, we calculated the AMOVA based on a distance matrix of the number of squared stepwise differences between pairs of samples.

Isolation by distance

To test for isolation by distance, we measured both direct Euclidean geographic and cost-path distances among study sites as described by Arora *et al.* (2010). Cost-path distances reflect the shortest distances between study sites that are unlikely to involve the crossing of a major river (for details, see Arora *et al.* 2010). As a measure of genetic distance, we used $\varphi'_{ST}/(1 - \varphi'_{ST})$ for Y-chromosomal as well as for mitochondrial genetic diversity, as φ'_{ST} corrects for varying within-population diversity (Hedrick 2005). Our results are qualitatively robust in respect to which measure of genetic distance (φ'_{ST}, φ_{ST} or R_{ST}) was used, and the quantitative differences are small (data not shown). Tests of matrix correlations (Mantel 1967) between genetic distances and the natural logarithm of geographic distances (Rousset 1997) were carried out with 10 000 randomizations using the R package VEGAN (Dixon 2003), as were partial Mantel tests to control if cost-path distance explained additional variation after correcting for direct geographic distance. To test whether the slopes of the regression of genetic distance on geographic distance differed between Y-chromosomal and mitochondrial markers, we used the R script MantelPiece (Postma 2011). MantelPiece is an extension of standard Mantel tests, as it randomizes the geographic distance matrix and then estimates the slopes for the randomized geographic distance matrix and the two genetic distance matrices, as well as the difference between these two slopes. Doing this 10,000 times generates a distribution for the null hypothesis of equal

slopes, from which a *P*-value for the observed difference in slopes can be calculated.

The linear regression slopes g of isolation-by-distance plots (pairwise genetic differentiation measure, here $\varphi'_{ST}/(1 - \varphi'_{ST})$, plotted against the natural logarithm of distance) can be used to calculate the dispersal distance σ, if effective population density D_e is known (Rousset 1997; Watts *et al.* 2007): $g = 1/(4\pi D_e \sigma^2)$. For orang-utans, however, there are no good data on effective population size or density, especially not for males and females separately. Thus, in order to estimate male-to-female average dispersal distances relative to each other, we combined and rearranged the equations for male and female dispersal to yield a ratio of male to female dispersal distance conditional on the ratio of female to male effective population density or size (when considering ratios, population density and population size are equivalent, given that male and female orang-utans inhabit the same area): $\sigma_{♂}/\sigma_{♀} = \sqrt{(g_{♀}/g_{♂})} \times \sqrt{(D_{e,♀}/D_{e,♂})}$, where the term $\sqrt{(g_{♀}/g_{♂})}$ is calculated from the slopes of the linear regressions of mitochondrial ($g_{♀}$) and Y-chromosomal ($g_{♂}$) genetic distance on the natural logarithm of genetic distance, respectively.

Results

Haplotypic diversity of mitochondrial DNA sequences and Y-chromosomal data was compared for the same males from nine orang-utan study sites (Table 1, Fig. 2, Data S1, Data S2). Mean haplotypic diversity over the seven Bornean populations was significantly higher for Y-chromosomal (0.82) than for mitochondrial (0.45) markers ($P < 0.01$, Wilcoxon rank sum test). Except of North Kinabatangan, the diversity of Y-chromosomal markers was higher than that of mitochondrial HVRI sequences (Fig. 2).

Median-joining networks of complete Y-chromosomal and mitochondrial haplotypes were constructed to illustrate the genetic variation and clustering due to geography. There was a clear separation of haplotypes from Bornean and Sumatran orang-utans, for both Y-chromosomal and mitochondrial data (Fig. 1). Qualitative assessment of both networks revealed that mitochondrial haplotypes showed much more geographic structuring than Y-chromosomal haplotypes (Fig. 1), to a degree that mitochondrial haplotypes from the same study site generally grouped closely together in the network, whereas Y-chromosomal haplotypes did not. Mitochondrial DNA haplotype sharing on Borneo occurs only at the three northernmost sites and between Sabangau and Gunung Palung. φ_{ST} indices based on mtDNA markers were significantly different between most pairs of populations (Table 2). In contrast, the network based on Y-chromosomal data showed no clear geographic clustering of haplotypes and is much more heterogeneous than the mtDNA network (Fig. 1). The pairwise R_{ST} values between populations were lower than for mtDNA, but still significant for most population pairs (Table 2).

Average genetic differentiation across Bornean study sites was calculated using three different statistics (R_{ST}, φ_{ST} and φ'_{ST}; Fig. 3). All differentiation measures for mitochondrial data were significantly higher than all differentiation measures for Y-linked data, as assessed by their nonoverlapping 95% confidence intervals. It should be noted that φ_{ST} for mitochondrial sequences is 2.1-fold higher than R_{ST} for Y-microsatellites, which is very similar to the ratio of φ'_{ST} for *Y*-linked and mitochondrial data (1.9-fold), implying that possible violations of mutation models affect the results only very marginally.

We further investigated the genetic differentiation among Bornean orang-utan males by conducting a hier-

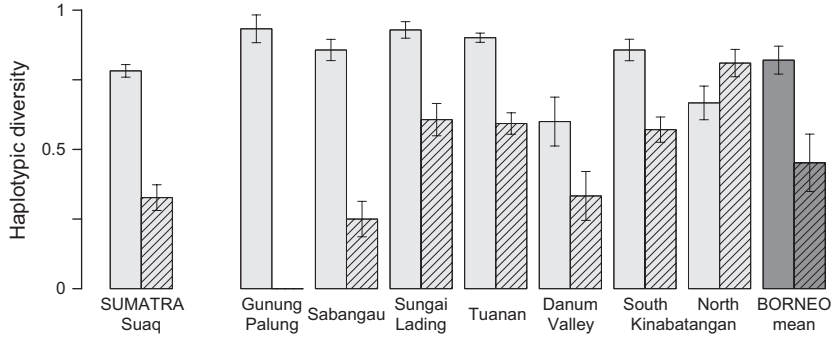


Fig. 2 Haplotypic diversity corrected for sample size is plotted for eight orang-utan populations and their mean over the Bornean populations. Plain bars represent Y-chromosomal haplotypic diversity, thatched bars represent mitochondrial diversity, and whiskers represent standard errors. Gunung Palung had only one mitochondrial haplotype, and hence, mitochondrial haplotypic diversity of 0. Ketambe is not plotted due to its low sample size (Table 1). The mean mitochondrial and Y-chromosomal values over all Bornean populations were significantly different ($P = 0.007$, Wilcoxon rank sum test).

Table 2 Genetic differentiation measures φ_{ST} for mitochondrial HVR-I sequences (above diagonal) and R_{ST} for Y-specific microsatellites (below diagonal) for pairwise comparisons among Bornean study sites

	DV	GP	NK	SK	SL	TU	WS
Danum Valley		0.98*	0.68*	0.23	0.85*	0.52*	0.96*
Gunung Palung	0.33*		0.90*	0.96*	0.90*	0.72*	−0.04
North Kinabatangan	0.46*	0.34*		0.68*	0.84*	0.63*	0.90*
South Kinabatangan	−0.15	0.43*	0.54*		0.83*	0. 53*	0.95*
Sungai Lading	0.34*	0.42*	0.28*	0.40*		0.66*	0.90*
Tuanan	0.27*	0.50*	0.47*	0.33*	0.07		0.74*
West Sabangau	0.41*	0.64*	0.50*	0.45*	0.07	0.35*	

Asterisks denote significance at $P < 0.05$.

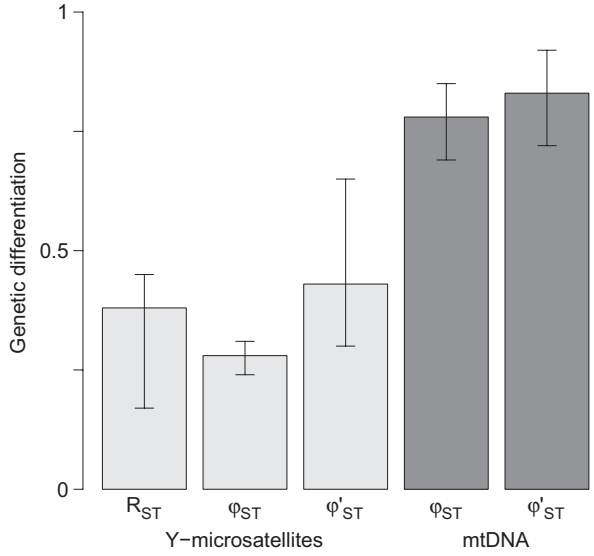


Fig. 3 Average genetic differentiation across Borneo for Y-chromosomal microsatellites (light grey bars) and mitochondrial HVRI sequences (dark grey bars). Whiskers represent 95% confidence intervals.

archical AMOVA with regional grouping of study sites less than 100 km apart. Most of the Y-chromosomal variation (61.9%; Table 3) was contained within study sites, whereas only a fifth of the mitochondrial variation lay within study sites (21.0%). Consequently, variation among study sites within geographic regions was much greater for mitochondrial (64.9%) than for *Y*-linked data (34.6%). For neither Y-linked nor mitochondrial data did region explain a significant part of genetic variation, although this value was much greater for mitochondrial data (Table 3).

To investigate the patterns of isolation by distance, we plotted pairwise values of $\varphi'_{ST}/(1 - \varphi'_{ST})$ for Y-linked and mitochondrial data against the natural logarithm of geographic distance among study sites (Fig. 4). Genetic distance increased significantly with geographic distance for both mitochondrial data (Mantel test, $P = 0.04$) and Y-linked data (Mantel test, $P = 0.01$). Although the slope of isolation by distance was more than one order of magnitude steeper for mitochondrial (=2.70) than for *Y*-linked data (=0.16), it was not significant (Postma's paired Mantel test, $P = 0.12$). Partial Mantel tests showed that cost-path distances treating major rivers as dispersal barriers did not explain a significant part of the remaining variation after correction for direct geographic distance, neither for mitochondrial ($P = 0.09$) nor for *Y*-linked data ($P = 0.13$).

We used the slopes of the isolation-by-distance plot (Fig. 4) to calculate the ratios of male to female dispersal distance conditional on ratios of female to male effective population sizes. Although caution needs to be exerted given the fairly large scatter of mtDNA data points, it appears that average male dispersal distances are at least 4.1-fold larger than those of females.

Discussion

Genetic structure among populations was consistently greater for maternally transmitted mitochondrial DNA, whereas the paternally transmitted Y-chromosomal markers showed higher variation within populations. Thus, all our data show strongly male-biased long-distance dispersal in orang-utans. Comparing the diversity of maternally and paternally transmitted genetic markers is a good approximation for which sex contributes more to gene flow among populations, although haplotypic diversity also depends on the number and mutation rates of the loci included. Haplotypic diversity of orang-utan Y chromosomes within single populations was significantly greater than that of mitochondrial sequences, pointing towards females being the sex that drives differentiation between populations and males accounting for the bulk of gene flow. In support of this logic, opposite patterns of haplotypic diversity were found in the clearly patrilocal bonobos (Eriksson *et al.* 2006) and chimpanzees (Langergraber *et al.* 2007b).

Table 3 Results of AMOVA for Y-chromosomal microsatellite data (based on the distance measure R_{ST}) and mitochondrial sequence data (based on the distance measure φ_{ST}) with grouping of study sites into three geographic regions (region 1: Gunung Palung; region 2: Sabangau, Sungai Lading and Tuanan; region 3: Danum Valley, South and North Kinabatangan)

		Y-linked microsatellites: R_{ST}				Mitochondrial DNA (HVRI): φ_{ST}			
Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation	Permutation probability	Sum of squares	Variance components	Percentage of variation	Permutation probability
Among regions	2	59.35	0.278	3.5	0.35	59.82	0.581	14.1	0.19
Among study sites within regions	4	109.50	2.761	34.6	0.00	90.74	2.685	64.9	0.00
Within study sites	49	242.13	4.941	61.9	0.00	42.55	0.868	21.0	0.00
Total	55	410.98	7.941			193.11	4.134		

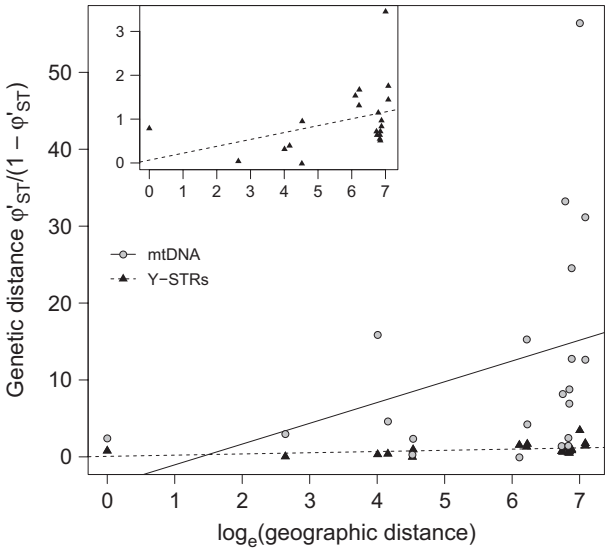


Fig. 4 Plot of genetic distance $\varphi'_{ST}/(1 - \varphi'_{ST})$ versus logarithmic direct geographic distance for pairwise comparisons among study sites. The inset graph shows the same Y-chromosomal data as in the large plot, but at a more revealing scale. Genetic distance increases significantly with geographic distance between study sites for mitochondrial sequences (Mantel test, $P = 0.04$), as well as Y-linked microsatellites (Mantel test, $P = 0.01$).

The grouping patterns of haplotypes in median-joining networks revealed geographic structuring differing for maternally and paternally transmitted markers. The network for mitochondrial haplotypes showed clear grouping according to the geographic origin, as observed in a previous study (Arora *et al.* 2010), whereas networks of Y-chromosomal haplotypes did not. The same pattern was found in haplotype networks that included both male and female samples and spanned most of their distributional range, including most known Sumatran populations (Nater *et al.* 2011). It is important to note that most variation in the Y-chromosomal data stems from microsatellite loci. Due to higher mutation rates compared to HVR-I sequences, homoplasies may be more problematic for microsatellite data. Median-joining networks constructed from Y-chromosomal SNPs and a subset of microsatellite loci with low variability still separated Bornean and Sumatran orang-utans, but showed no indications of geographic grouping of similar haplotypes and were less resolved (data not shown). Thus, our study does not seem to be affected by homoplasies.

Even for geographically close populations, genetic differentiation was pronounced and mostly significant for mitochondrial data, but less often so for Y-chromosomal data. As most compared populations had at least one major river between them, this high differentiation is best explained by rivers acting as dispersal barriers, as had been found, for example, for the Kinabatangan River in northern Borneo (Goossens *et al.* 2005; Jalil *et al.* 2008) and on Borneo in general (Warren *et al.* 2001; Arora *et al.* 2010).

Average differentiation among study sites is expected to be greater for the markers transmitted through the more philopatric sex, as reduced migration and thus lower gene flow cannot sufficiently counteract the differentiating effects of genetic drift or local adaptation. Despite some difficulties with some measures of genetic structure, in case of the orang-utan data set used in this study, the choice of differentiation measure only quantitatively affected the results and even that only very slightly. Comparing R_{ST} for Y-linked loci with φ_{ST} for mitochondrial data or φ'_{ST} for both data sets both showed that the differentiation values for mitochondrial data are about twice as high as for Y-chromosomal data. Opposite patterns were found in bonobos (Eriksson *et al.* 2006) and chimpanzees (Langergraber *et al.* 2007b), which both show clearly female-biased dispersal. The results for gorillas are more complex because of the harem structure of the gorilla breeding system. Douadi *et al.* (2007) found no genetic structure among sites (comprising multiple breeding groups) for Y-chromosomal

variation, but did for mitochondrial sequences, indicating that males disperse further than females.

As expected under male-biased dispersal, we found that genetic distance increased nearly 17 times faster with geographic distance for mitochondrial data than for Y-linked markers, that is, faster for the marker system transmitted through the more philopatric sex. After correction for direct geographic distance, cost-path distance around large rivers did not explain a significant part of the remaining variation for either marker system. Arora *et al.* (2010) carried out a similar analysis and found that cost-path distances after accounting for direct distances explained a significant part of the variation for mitochondrial data, but not autosomal microsatellites. In contrast to our study, their analyses included samples from both male and female adult orang-utans, which may be the reason for this discrepancy, as rivers appear to be a less effective barrier for dispersing males than for more philopatric females. This explanation does not imply that males actually cross rivers at points where females would not, but rather that males move considerably further, which makes it more likely that males or their descendants may reach a point where the rivers become crossable.

We estimated average dispersal distance ratios for males and females. For all female-biased effective population size ratios, male dispersal distance is estimated to be at least 4.1-fold larger than that of females. Although we know that females tend to settle adjacent to their mother's home range (van Noordwijk *et al.* 2012), quantitative evidence for average dispersal distance of any sex is difficult to gather in orang-utans and thus lacking, precluding absolute quantification of average dispersal distances. Our relative quantification relies on the ratio of female to male effective population size. In most animals, due to sexual selection and highly female-biased parental care for offspring, male lifetime reproductive success varies more than female reproductive success (Trivers 1972), as shown for, for example, red deer (*Cervus elaphus*; Kruuk *et al.* 1999) and tiger (*Panthera tigris*; Smith & McDougal 1991). The only exceptions seem to be species where male parental care is a limiting resource for females (Trivers 1972). Although data on lifetime reproductive success are lacking in orang-utans, there is no obvious reason to assume that they would be an exception. Males do not offer parental care and sexual dimorphism in body size and weapons (*i.e.* canines) is pronounced (Plavcan & van Schaik 1997). Furthermore, copulations with resisting females occur (Utami Atmoko *et al.* 2009). This all suggests that access to males and their parental care is not a limiting resource to females. Orang-utan males are thus expected to show stronger reproductive skew than females and that variation in lifetime reproductive success is higher in males, thus reducing male effective population size (Halliburton 2004). An additional factor that affects effective population size is different average generation length among males and females (*i.e.* longer generation length increases effective population size; Caballero 1994). Although data on age-dependent reproductive success in wild orang-utans are lacking, it is conceivable that orang-utan males may have a longer average generation time than females, as the older flanged males possibly sire more offspring than the younger unflanged males, although both morphs acquire paternities (Utami *et al.* 2002; Goossens *et al.* 2006b). Furthermore, males reach sexual maturity slightly later than females (reviewed in Delgado & van Schaik 2000). Thus, although the difference may be reduced due to generation length differences, we still expect a higher female than male effective population size.

It is important to bear in mind that the estimation of relative dispersal distances relies crucially on the assumptions of the underlying model, one of them being that populations are in migration–drift equilibrium (Rousset 1997; Whitlock & McCauley 1999); especially in the case of recently diverged orang-utans (Arora *et al.* 2010) with low female dispersal rates (Arora 2011) and expected higher female effective population size, reaching migration–drift equilibrium may take longer for mtDNA markers than for Y chromosomes (Whitlock 1992), which would affect dispersal distance ratio estimates. Yet, time differences in reaching equilibrium for Y chromosomes and mtDNA would partly be a consequence of different male and female dispersal patterns. Furthermore, the applied model assumes that the whole island of Borneo is continuous orang-utan habitat without significant barriers. This assumption may be violated as there is a strong indication that major rivers do affect orang-utan dispersal (Arora *et al.* 2010). Furthermore, habitat fragmentation on Borneo also poses a, albeit very recent, problem for dispersal, and thus could quantitatively affect the relative dispersal distance estimation, although only if male and female dispersal patterns are differentially affected.

The conclusion that orang-utan dispersal is male-biased is in good agreement with behavioural observations (Rijksen 1978; van Schaik & van Hooff 1996; Delgado & van Schaik 2000; van Noordwijk *et al.* 2012). Conflicting findings from studies using autosomal microsatellites (Utami *et al.* 2002; Goossens *et al.* 2006b; Morrogh-Bernard *et al.* 2011) can be explained as a result of the peculiarities of each site as well as the problems associated with the interpretation of average pairwise relatedness coefficients without the incorporation of spatial, behavioural or other genetic data (Arora *et al.*, submitted). Our findings, however, are the first to

unequivocally show that long-distance male-biased dispersal is the norm in orang-utan species and is responsible for limiting the extent of genetic differentiation among orang-utan populations.

Male-biased dispersal appears to be the norm in most mammals (Greenwood 1980), probably including gorillas (Douadi *et al.* 2007). The opposite pattern is found in chimpanzees (Langergraber *et al.* 2007b) and bonobos (Eriksson *et al.* 2006) and, according to behavioural studies, appears to be the most common pattern in humans (reviewed in Burton *et al.* 1996; Hill *et al.* 2011). There is no uniform picture in human studies using genetics (Seielstad *et al.* 1998; Stoneking 1998; Wilder *et al.* 2004), probably due to confounding effects of effective population size, whose sex bias differs among populations (Segurel *et al.* 2008). In chimpanzees, male coalitions are an important factor for male reproductive success (Goodall 1986), making it advantageous for males to stay close to kin, as maternal kin are thought to be more reliable coalition partners (Langergraber *et al.* 2007a; Mitani 2009). Orang-utans are semi-solitary (Galdikas 1985), and male coalitions do not play any role (van Schaik & van Hooff 1996). Hence, the advantages for females to stay in a familiar environment and to live close to maternal kin probably outweigh the corresponding benefits for males (Lawson Handley & Perrin 2007), as shown in a recent study by van Noordwijk *et al.* (2012). The same conclusion holds for gorillas, which usually live in groups of one adult male and several females with their offspring (Harcourt & Stewart 2007). Taken together, in the great ape lineage, life in multimale groups seems to be associated with the evolution of male philopatry. The benefits for males could come from profiting from established social bonds, be it with kin or nonkin.

We showed here that male dispersal, which can be long-distance, has historically largely been responsible for gene flow among orang-utan populations. Orang-utans occur on the islands of Borneo (current estimates of about 50 000 animals) and Sumatra (6500 animals) (Wich *et al.* 2008) and are rated as endangered and critically endangered, respectively (IUCN 2012). The recent drastic decline in population size (Goossens *et al.* 2006a) is mostly attributed to habitat loss (IUCN 2012). Thus, extending our knowledge with regard to dispersal patterns also has ramifications for the conservation of the genus. Loss of male dispersal would significantly increase the effects of genetic drift due to small local population sizes and extreme female philopatry. In order to conserve the natural genetic structure of orang-utan populations, it is therefore crucial to ensure that males are able to move between suitable habitat patches, even if they are far apart. An effective immigration rate of one to two animals per generation is considered to be the minimum to reduce the negative fitness effects of inbreeding depression (Lopez *et al.* 2009). Thus, it is highly desirable that all extant orang-utan populations are connected through gene flow. If orang-utans are to be efficiently protected, a sufficient network of high-quality natural forest and dispersal corridors has to be restored across Borneo and Sumatra to allow orang-utans to disperse naturally.

Acknowledgements

We are grateful to Maria van Noordwijk, Erik Postma, Lukas Keller and two anonymous reviewers for invaluable comments on earlier drafts of the manuscript, as well as Erik Willems for providing cost-path distances and the distribution map in Fig. 1. We very much appreciate the help of numerous individuals for collecting samples at the various field sites. We are indebted to Joko Pamungkas and Dyah Perwitasari-Farajallah from the Indonesian Primate Centre, University of Bogor for their invaluable help with permits. We thank the Indonesian Institute of Sciences (LIPI), the Indonesian State Ministry for Research and Technology (RISTEK) and the Sabah Wildlife Department for granting permission to undertake this research. Samples were exported under the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) from Indonesia, Malaysia and the U.K. to Switzerland with permits 07279/IV/SATS-LN/2009, 00961/IV/SATS-LN/2007, 4645, 06968/IV/SATS-LN/2005 and 290569/01. Funding for this study was obtained from the Swiss National Science Foundation (grant 31003A-116848 to MK and CvS), A. H. Schultz-Stiftung, Messerli Stiftung, Claraz Schenkung, Leakey Foundation and the Wildlife Conservation Society.

References

Arora N, Nater A, van Schaik CP *et al.* (2010) Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (*Pongo pygmaeus*). *Proceedings of the National Academy of Sciences*, **107**, 21376–21381.

Arora N (2011) Orangutan genes in space and time: the impact of evolutionary processes of diversification on Bornean orangutans. PhD thesis, University of Zurich, Switzerland.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Burton ML, Moore CC, Whiting JWM, Romney AK (1996) Regions based on social structure. *Current Anthropology*, **37**, 87–123.

Caballero A (1994) Developments in the prediction of effective population size. *Heredity*, **73**, 657–679.

Delgado RA, van Schaik CP (2000) The behavioral ecology and conservation of the orang-utan (*Pongo pygmaeus*): a tale of two islands. *Evolutionary Anthropology*, **9**, 201–218.

Dixon P (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Douadi MI, Gatti S, Levrero F *et al.* (2007) Sex-biased dispersal in western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology*, **16**, 2247–2259.

Eriksson J, Siedel H, Lukas D *et al.* (2006) Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (*Pan paniscus*). *Molecular Ecology*, **15**, 939–949.

Erler A, Stoneking M, Kayser M (2004) Development of Y-chromosomal microsatellite markers for nonhuman primates. *Molecular Ecology*, **13**, 2921–2930.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes – Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Galdikas BMF (1985) Adult male sociality and reproductive tactics among orangutans at Tanjung Puting. *Folia Primatologica*, **45**, 9–24.

Goodall J (1986) *The Chimpanzees of Gombe: Patterns of Behavior*. Harvard University Press, Cambridge, UK.

Goossens B, Chikhi L, Jalil MF *et al.* (2005) Patterns of genetic diversity and migration in increasingly fragmented and declining orang-utan (*Pongo pygmaeus*) populations from Sabah, Malaysia. *Molecular Ecology*, **14**, 441–456.

Goossens B, Chikhi L, Ancrenaz M *et al.* (2006a) Genetic signature of anthropogenic population collapse in orang-utans. *Plos Biology*, **4**, 285–291.

Goossens B, Setchell JM, James SS *et al.* (2006b) Philopatry and reproductive success in Bornean orang-utans (*Pongo pygmaeus*). *Molecular Ecology*, **15**, 2577–2588.

Greenwood PJ (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, **28**, 1140–1162.

Greminger MP, Krützen M, Schelling C, Pienkowska-Schelling A, Wandeler P (2010) The quest for Y chromosome-specific markers – methodological strategies for mammalian non-model organisms. *Molecular Ecology Resources*, **10**, 409–420.

Halliburton R (2004) *Introduction to Population Genetics*. Pearson Prentice Hall, Upper Saddle River.

Harcourt AH, Stewart KJ (2007) Gorilla society: what we know and don't know. *Evolutionary Anthropology*, **16**, 147–158.

Harcourt AH, Stewart KS, Fossey D (1976) Male emigration and female transfer in wild mountain gorilla. *Nature*, **263**, 226–227.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hill KR, Walker RS, Božičević M *et al.* (2011) Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science*, **331**, 1286–1289.

Hohmann G (2001) Association and social interactions between strangers and residents in bonobos (*Pan paniscus*). *Primates*, **42**, 91–99.

IUCN (2012) IUCN Red List of threatened species. http://www.iucnredlist.org.

Jalil MF, Cable J, Inyor JS *et al.* (2008) Riverine effects on mitochondrial structure of Bornean orang-utans (*Pongo pygmaeus*) at two spatial scales. *Molecular Ecology*, **17**, 2898–2909.

Jost L (2008) G(ST) and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Jost L (2009) D vs. G(ST): response to Heller and Siegismund (2009) and Ryman and Leimar (2009). *Molecular Ecology*, **18**, 2088–2091.

Kayser M, Kittler R, Erler A *et al.* (2004) A comprehensive survey of human Y-chromosomal microsatellites. *American Journal of Human Genetics*, **74**, 1183–1197.

Kronholm I, Loudet O, de Meaux J (2010) Influence of mutation rate on estimators of genetic differentiation – lessons from *Arabidopsis thaliana*. *Bmc Genetics*, **11**, 1–14.

Kruuk LEB, Clutton-Brock TH, Rose KE, Guinness FE (1999) Early determinants of lifetime reproductive success differ between the sexes in red deer. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **266**, 1655–1661.

Langergraber KE, Mitani JC, Vigilant L (2007a) The limited impact of kinship on cooperation in wild chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 7786–7790.

Langergraber KE, Siedel H, Mitani JC *et al.* (2007b) The genetic signature of sex-biased migration in patrilocal chimpanzees and humans. *PLoS One*, **2**, e973.

Lawson Handley LJ, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. *Molecular Ecology*, **16**, 1559–1578.

Lopez S, Rousset F, Shaw FH, Shaw RG, Ronce O (2009) Joint effects of inbreeding and local adaptation on the evolution of genetic load after fragmentation. *Conservation Biology*, **23**, 1618–1627.

Lukas D, Reynolds V, Boesch C, Vigilant L (2005) To what extent does living in a group mean living with kin? *Molecular Ecology*, **14**, 2181–2196.

Luo SJ, Johnson WF, David VA *et al.* (2007) Development of Y chromosome intraspecific polymorphic markers in the Felidae. *Journal of Heredity*, **98**, 400–413.

Mantel N (1967) Detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–20.

Masters BS, Johnson LS, Johnson BGP *et al.* (2010) Evidence for heterozygote instability in microsatellite loci in house wrens. *Biology Letters*, **7**, 127–130.

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.

Mitani JC (2009) Male chimpanzees form enduring and equitable social bonds. *Animal Behaviour*, **77**, 633–640.

Morrogh-Bernard HC, Morf NV, Chivers DJ, Krützen M (2011) Dispersal patterns of orang-utans (*Pongo* spp.) in a Bornean peat-swamp forest. *International Journal of Primatology*, **32**, 362–376.

Nater A, Nietlisbach P, Arora N *et al.* (2011) Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant orangutans (genus: *Pongo*). *Molecular Biology and Evolution*, **28**, 2275–2288.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.

Nietlisbach P, Nater A, Greminger M, Arora N, Krützen M (2010) A multiplex-system to target 16 male-specific and 15 autosomal genetic markers for orang-utans (genus: *Pongo*). *Conservation Genetics Resources*, **2**, 153–158.

Nishida T, Corp N, Hamai M *et al.* (2003) Demography, female life history, and reproductive profiles among the chimpanzees of Mahale. *American Journal of Primatology*, **59**, 99–121.

van Noordwijk MA, Arora N, Willems EP *et al.* (2012) Female philopatry and its social benefits among Bornean orangutans. *Behavioral Ecology and Sociobiology*, in press.

Nsubuga AM, Robbins MM, Roeder AD *et al.* (2004) Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology*, **13**, 2089–2094.

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nature Genetics*, **29**, 20–21.

Plavcan JM, van Schaik CP (1997) Intrasexual competition and body weight dimorphism in anthropoid primates. *American Journal of Physical Anthropology*, **103**, 37–68.

Postma E (2011) MantelPiece.R. http://www.erikpostma.net/resources.html.

Prugnolle F, de Meeus T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity*, **88**, 161–165.

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Raaum RL, Sterner KN, Noviello CM, Stewart CB, Disotell TR (2005) Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *Journal of Human Evolution*, **48**, 237–257.

Rijksen HD (1978) *A Field Study on Sumatran Orang Utans (Pongo pygmaeus abelii, Lesson 1827): Ecology, Behaviour and Conservation*. H. Veenman & Zonen, Wageningen, the Netherlands.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Ryman N, Leimar O (2009) G(ST) is still a useful measure of genetic differentiation – a comment on Jost's D. *Molecular Ecology*, **18**, 2084–2087.

van Schaik CP, van Hooff JARAM (1996) Toward an understanding of the orangutan's social system. In: *Great Ape Societies* (eds McGrew WC, Marchant LF and Nishida T), pp. 3–15. Cambridge University Press, Cambridge.

Segurel L, Martinez-Cruz B, Quintana-Murci L *et al.* (2008) Sex-specific genetic structure and social organization in central Asia: Insights from a multi-locus study. *Plos Genetics*, **4**, 14.

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nature Genetics*, **20**, 278–280.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Smith JLD, McDougal C (1991) The contribution of variance in lifetime reproduction to effective population size in tigers. *Conservation Biology*, **5**, 484–490.

Stoneking M (1998) Women on the move. *Nature Genetics*, **20**, 219–220.

Tajima F (1993) Measurement of DNA polymorphism. In: *Mechanisms of Molecular Evolution: Introduction to Molecular Paleopopulation Biology* (eds Takahata N and Clark AG), pp. 37–59. Japan Scientific Societies Press, Tokyo.

Trivers RL (1972) Parental investment and sexual selection. In: *Sexual Selection and the Descent of Man 1871–1971* (ed Campbell B). Aldine Publishing Company, Chicago. pp. 136–179.

Utami Atmoko SS, Mitra Setia T, Goossens B *et al.* (2009) Orang-utan mating behavior and strategies. In: *Orang-utans – Geographic Variation in Behavioral Ecology and Conservation* (eds Wich SA, Utami Atmoko SS, Mitra Setia S and van Schaik C). Oxford University Press, Oxford, UK. pp. 235–244.

Utami SS, Goossens B, Bruford MW, de Ruiter JR, van Hooff J (2002) Male bimaturism and reproductive success in Sumatran orang-utans. *Behavioral Ecology*, **13**, 643–652.

Warren KS, Verschoor EJ, Langenhuijzen S *et al.* (2001) Speciation and intrasubspecific variation of Bornean orang-utans, *Pongo pygmaeus pygmaeus*. *Molecular Biology and Evolution*, **18**, 472–480.

Watts DP (2000) Causes and consequences of variation in male mountain gorilla life histories and group membership. In: *Primate Males: Causes and Consequences of Variation in Group Composition* (ed Kappeler PM), pp. 169–179. Cambridge University Press, Cambridge, UK.

Watts PC, Rousset F, Saccheri IJ *et al.* (2007) Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion mercuriale*: Odonata: Zygoptera) populations: analysis of 'neighbourhood size' using a more precise estimator. *Molecular Ecology*, **16**, 737–751.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution*, **38**, 1358–1370.

Whitlock MC (1992) Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution*, **46**, 608–615.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity*, **82**, 117–125.

Wich SA, Meijaard E, Marshall AJ *et al.* (2008) Distribution and conservation status of the orang-utan (*Pongo* spp.) on Borneo and Sumatra: how many remain? *Oryx*, **42**, 329–339.

Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF (2004) Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nature Genetics*, **36**, 1122–1125.

PN, NA, AN, BG and MK apply genetic methods to study evolutionary and ecological questions pertaining to inbreeding, processes affecting genetic diversity, phylogeography, population structure and history, conservation of endangered mammals and the evolution of culture. CvS studies the evolution of human nature by comparative work and detailed field studies of orang-utans. BG is the director of a tropical field centre in Sabah, Malaysia.

Data accessibility

Data for mitochondrial HVR-I sequences are accessible in GenBank under accession numbers EU547193, EU547195, EU547198–EU547200, FR717918, FR717920–FR717923, FR717925–FR717927, FR717930, FR717933, FR717935, FR717936, FR717939, FR717940. The DNA alignment for all HRV-I sequences for every individual, as well as a spreadsheet containing the individual-

by-individual genotypes for the 18 Y-chromosomal markers, is given in the Supporting information.

Supporting information

Additional supporting information may be found in the online version of this article.

Data S1. DNA alignment for all HVR-I sequences.

Data S2. Individual genotypes for 18 Y-chromosomal markers.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (*Pongo pygmaeus*)

Natasha Arora[a,1], Alexander Nater[a], Carel P. van Schaik[a], Erik P. Willems[a], Maria A. van Noordwijk[a], Benoit Goossens[b,c], Nadja Morf[a], Meredith Bastian[d,e], Cheryl Knott[d], Helen Morrogh-Bernard[f], Noko Kuze[g], Tomoko Kanamori[h], Joko Pamungkas[i], Dyah Perwitasari-Farajallah[i], Ernst Verschoor[j], Kristin Warren[k], and Michael Krützen[a]

[a]Anthropological Institute and Museum, University of Zurich, 8057 Zurich, Switzerland; [b]Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, United Kingdom; [c]Sabah Wildlife Department, 88100 Kota Kinabalu, Malaysia; [d]Department of Anthropology, Boston University, Boston, MA 02215; [e]Philadelphia Zoo, Philadelphia, PA 19104; [f]Anatomy School, University of Cambridge, Cambridge CB2 3DY, United Kingdom; [g]Department of Zoology, Kyoto University, Kyoto 606-8502, Japan; [h]Tokyo Institute of Technology, Tokyo 152-8550 Japan; [i]Indonesian Primate Centre, Bogor Agricultural University, Bogor 16680, Indonesia; [j]Biomedical Primate Research Centre, Rijswijk, 2280, The Netherlands; and [k]School of Veterinary and Biomedical Sciences, Murdoch University, Murdoch 6150, Australia

Edited by Morris Goodman, School of Medicine, Wayne State University, Detroit, MI, and approved October 21, 2010 (received for review July 20, 2010)

Sundaland, a tropical hotspot of biodiversity comprising Borneo and Sumatra among other islands, the Malay Peninsula, and a shallow sea, has been subject to dramatic environmental processes. Thus, it presents an ideal opportunity to investigate the role of environmental mechanisms in shaping species distribution and diversity. We investigated the population structure and underlying mechanisms of an insular endemic, the Bornean orangutan (*Pongo pygmaeus*). Phylogenetic reconstructions based on mtDNA sequences from 211 wild orangutans covering the entire range of the species indicate an unexpectedly recent common ancestor of Bornean orangutans 176 ka (95% highest posterior density, 72–322 ka), pointing to a Pleistocene refugium. High mtDNA differentiation among populations and rare haplotype sharing is consistent with a pattern of strong female philopatry. This is corroborated by isolation by distance tests, which show a significant correlation between mtDNA divergence and distance and a strong effect of rivers as barriers for female movement. Both frequency-based and Bayesian clustering analyses using as many as 25 nuclear microsatellite loci revealed a significant separation among all populations, as well as a small degree of male-mediated gene flow. This study highlights the unique effects of environmental and biological features on the evolutionary history of Bornean orangutans, a highly endangered species particularly vulnerable to future climate and anthropogenic change as an insular endemic.

Asian great ape | genetic structure | radiation | geographical barriers | sociobehavioral barriers

Environmental mechanisms are some of the most important forces affecting the evolutionary history and current distribution of species. Such mechanisms have been invoked to explain genetic structure in many temperate European and North American species but with little focus on hotspots of biodiversity and endemism in the tropics (1), where the forces underlying patterns of genetic diversity and differentiation are especially intriguing.

The tropical Asian hotspot of Sundaland is remarkable in that it has been subject to dramatic geological and environmental changes (2, 3). This now partly submerged continental shelf encompasses the Malaysian peninsula, the islands of Borneo, Sumatra, Java, and possibly Palawan (2). It is a historically dynamic tectonic area that underwent notable landmass configuration changes (3). More recently, it has been severely affected by the Pleistocene climatic oscillations (4) of the Quaternary. Changes in sea levels resulted in the cyclical exposure of the continental shelf and the formation of land bridges between the islands (4, 5), allowing for species interchange with subsequent isolation (6). Moreover, climatic fluctuations were accompanied by vegetation changes (2, 7, 8), with shifts in the range and elevational distribution of rainforests. Thus, these changes led to habitat expansions or contractions, leading to new openings or barriers to gene flow. The Pleistocene was further punctuated by intense regional climatic and habitat changes through extraordinary volcanic eruptions, especially of Mount Toba (9, 10). Finally, Sundaland contains many interesting topographical features, including rivers, lakes, and mountains (5, 11, 12), that may have acted as barriers to dispersal for a number of species, adding yet another potential allopatric force.

The roles of these environmental forces in driving biotic diversity and endemism remain underexplored, particularly in Borneo, the world's second largest tropical island as well as the easternmost Sunda region abutting the Wallace line (13, 14). Its unusually high species endemism (14–16) suggests a combination of specialized ecological niches, refugia formation, and long periods of isolation.

Among the species endemic to the island are the Bornean orangutans (*Pongo pygmaeus*). This rainforest canopy-bound species with an unusually slow life history is characterized by a rich spectrum of genetic, morphological, and cultural variation (17–19). Fossils indicate a much wider distribution of orangutans during the Pleistocene extending from Southern China and Vietnam to Java (11, 18), but orangutans are currently only found, as distinct species, in Borneo (*P. pygmaeus*) and Sumatra (*Pongo abelii*). The ancestors of orangutans therefore probably migrated from the mainland to Sumatra and from there to Borneo (12), yet it remains unclear when and how these colonization events took place.

It is also unclear how the exceptional environmental features of Sundaland, combined with the characteristic behavioral and ecological traits of orangutans, have shaped their phylogeography. For instance, isolation in refugia or through riverine barriers have been described as important forces underlying the genetic structure of some of the African great apes (20–22), yet the evolutionary history of orangutans remains unresolved. First, the high genetic differentiation between Bornean and Sumatran orangutans (17, 23) is intriguing given recurrent land bridge formation between the islands during the Pleistocene glacial periods (5). Second, within Borneo, arguments for a stable distribution since colonization (24) clash with that of a bottleneck possibly associated

Author contributions: N.A., C.P.v.S., and M.K. designed research; N.A., A.N., B.G., and N.M. performed research; M.A.v.N., B.G., M.B., C.K., H.M.-B., N.K., T.K., J.P., D.P.-F., E.V., and K.W. coordinated sample collection; N.A. and E.P.W. analyzed data; and N.A., C.P.v.S., and M.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: n.arora@aim.uzh.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1010169107/-/DCSupplemental.

with the last eruption of Mount Toba (25). Third, the three Bornean subspecies (*P. p. pygmaeus, P. p. wurmbii, P. p. morio*), described on the basis of morphological characteristics (26), show unexplained genetic substructuring (17). Fourth, as for geographical features, the marked role of rivers as dispersal barriers has been highlighted in the study of populations in Sabah (27, 28), but it remains to be seen whether other rivers have had similar vicariant effects. Thus, the relative importance of the Pleistocene sea level and vegetation changes, Toba eruptions, and rivers as dispersal barriers, against the background of regular dispersal behavior of orangutans, remains unknown.

These questions also acquire special relevance today from a conservation perspective, in the light of ongoing habitat conversion (29) and predicted future climate change (30, 31), particularly for insular endemics and highly endangered species such as orangutans.

We recently obtained noninvasively collected wild Bornean orangutan samples from seven long-term study sites, as well as other localities, thus encompassing most of the species' range. Capitalizing on the most extensive sample size to date, we provide genetic evidence for a recent radiation of Bornean populations within the Middle to Late Pleistocene. We further illustrate the role of rivers and sex-biased dispersal in generating the marked population structure of the largest arboreal primate.

Results

mtDNA Analyses. We generated a phylogenetic tree for the mitochondrial (mtDNA) haplotypes from 211 individuals distributed throughout 10 sampling sites in Borneo (Fig. 1*B*), as well as six Sumatran individuals. The tree (Fig. 1*A*) shows a monophyletic Bornean clade with a surprisingly recent mean coalescence date of 176 ka (95% highest posterior density, 72–322 ka), contrasting with a much older estimate from a previous study (17). The phylogenetic tree and divergence estimate further illustrate the deeper coalescence of Bornean and Sumatran haplotypes (mean, 3.6 Ma; 95% highest posterior density, 2.3–5.0 Ma). Given the recurrent formation of potential connections between the islands, these findings point to an unexpectedly recent and single origin for current Bornean populations. Furthermore, the Bornean subspecies, as currently recognized on the basis of morphological

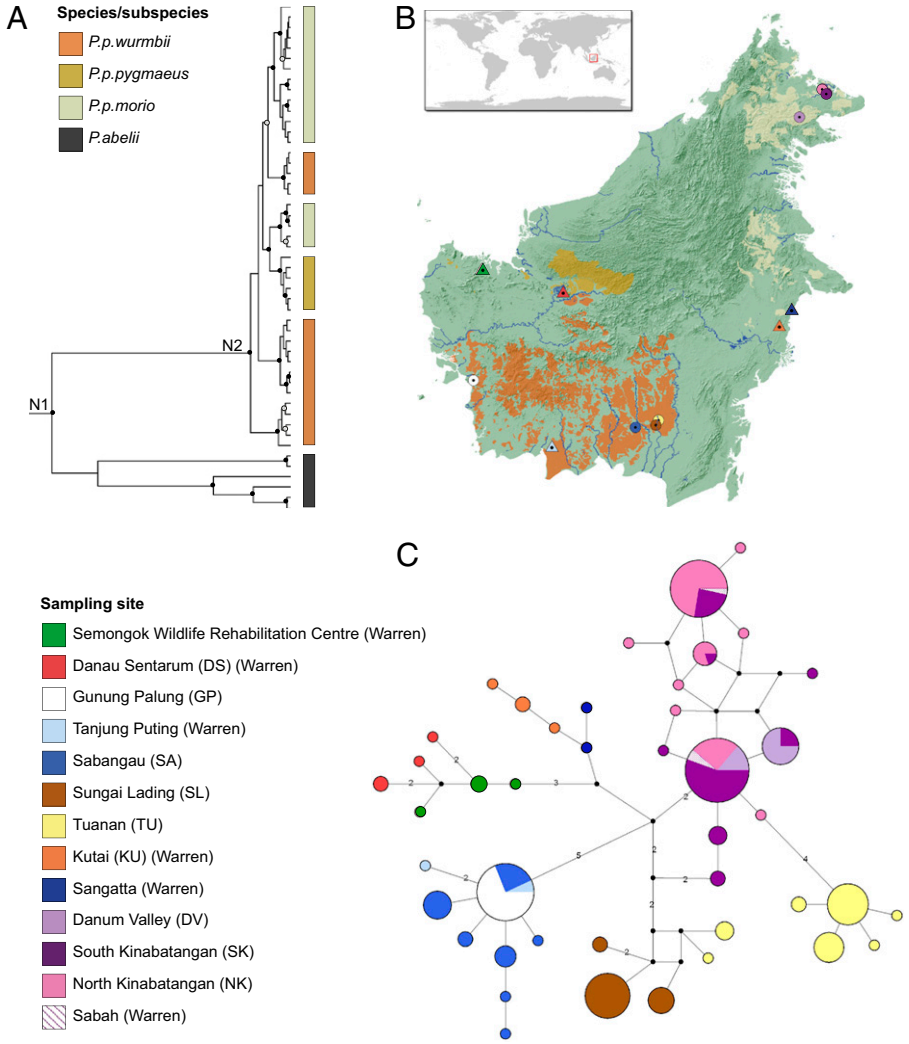


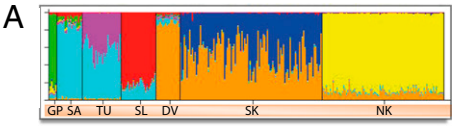
Fig. 1. Phylogenetic reconstruction and sampling sites of Bornean orangutans. (*A*) Bayesian phylogenetic tree of Bornean and Sumatran mtDNA haplotypes. Circles show posterior probabilities (>0.5, open circles; >0.75, black circles). Colored bars next to tips indicate species/subspecies designation. (*B*) Map of Borneo with location of sampling sites. Triangles correspond to sites for which only mtDNA data are available, circles correspond to sites for which additionally microsatellite data are available. Colored ranges on the map represent subspecies. (*C*) Median joining network of Bornean mtDNA HVRI haplotypes. Mutational steps are one unless indicated by the numbers. Two haplotypes from TU more closely related to those from SL are exclusively found in males. Sites with resequenced data from Warren et al. (17) are indicated in parentheses.

characteristics, are not reciprocally monophyletic, and should therefore be reconsidered.

The surprisingly recent radiation of a single Bornean lineage calls for a more detailed exploration of Bornean phylogeography. We generated an mtDNA phylogenetic network (Fig. 1*C*), more appropriate for population level studies than phylogenetic trees as they do not force possible ancestral haplotypes to the tips (32, 33). The network revealed seven main star-like geographical clusters, reflecting considerable structuring within the different subspecies. These seven clusters were further supported by a spatial analysis of molecular variance (SAMOVA), which defines groups of populations that are "geographically homogeneous and maximally differentiated from each other" (34). The analysis indicated that among-group variance asymptotes at 79.27% (F_{CT} = 0.793, $P < 0.01$) with seven groups of populations. The grouping corresponds to an almost complete separation of all sampled sites except for: (*i*) Danum Valley (DV), which clusters with South Kinabatangan (SK), a site in close proximity (approximately 90 km) not separated by geographical barriers (Fig. 1*B*); and (*ii*) Gunung Palung (GP), clustering with Sabangau (SA), a site with which it shares its only haplotype. Our results point to strong interpopulation differentiation for mtDNA, as corroborated by the high and significant Φ_{ST} values for all 36 population pairs (Fig. 2*B*). The exceptions are three lower, albeit still significant, Φ_{ST} values between the sites that share haplotypes. Given the heavy reliance of Φ_{ST} and other classic moment-based estimators on intrapopulation diversity (35), we also computed population average pairwise differences (Table S1). We found generally higher levels of diversity between populations than within, providing additional support for interpopulation differentiation.

Microsatellite Analyses. We also examined differentiation patterns using nuclear loci, which are biparentally inherited and therefore representative of both male and female histories, for the seven sites for which we could generate microsatellite genotypes. Both cluster analyses with Structure and significant pairwise population F_{ST} values indicate strong structuring of these sites (Fig. 2), particularly when separated by rivers (Fig. 1*B*). The structure runs for all seven sites using 12 microsatellite loci (dataset II, Fig. 2*A*) yielded the highest probability runs for K = 7 [Log likelihood (LnL), −9,619.88], partitioning each of the sites as a distinct cluster. Likewise, a more detailed analysis for the five sites for which 25 microsatellite loci were available (dataset I) also led to each one being inferred as a separate cluster (Fig. S1). Generally, high pairwise F_{ST} and level of structuring of populations is congruent with our mtDNA results. However, the cluster analyses using nuclear loci indicate some heterogeneity within populations. As haplotype sharing is rare among populations exchanging migrants, the low levels of gene flow are most likely male-mediated.

We investigated the signature of sex-specific demographic processes more directly by comparing isolation by distance patterns for the nuclear and mtDNA loci. The Mantel test for the relationship between genetic and Euclidean geographical distance yielded a significant and positive correlation for both the nuclear markers and mtDNA (F_{ST}, $r = 0.415$, $P < 0.05$; Φ_{ST}, $r = 0.357$, $P < 0.05$). We also explored the effect of rivers in a partial Mantel test of the association between genetic and cost path distances while controlling for Euclidean distance. Results were significant for the mtDNA ($P < 0.01$; $r = 0.403$) but not the nuclear markers ($P = 0.633$; $r = -0.096$). It is noteworthy, however, that for the mtDNA, only three of the 36 population pairs studied have low Φ_{ST} values (<0.6). Therefore, most populations are highly differentiated from each other despite the short geographical distances between them.



B

	DS	KU	GP	SA	SL	TU	DV	SK	NK
DS		-	-	-	-	-	-	-	-
KU	0.790*		-	-	-	-	-	-	-
GP	0.962	0.986		0.143**	0.160	0.163	0.209**	0.131	0.125
SA	0.880	0.882	0.159		0.110	0.049	0.152	0.083	0.069
SL	0.881	0.874	0.934	0.895		0.084	0.190	0.109	0.121
TU	0.758	0.680	0.827	0.801	0.779		0.117	0.072	0.090
DV	0.910	0.914	0.970	0.893	0.886	0.658		0.036	0.112
SK	0.812	0.754	0.845	0.810	0.814	0.628	0.163**		0.040
NK	0.879	0.854	0.896	0.859	0.879	0.748	0.689	0.409	

Fig. 2. Population structure based on nuclear microsatellite markers. (*A*) Structure run for the seven study sites with 12 microsatellite marker data (dataset II) at K = 7 (LnL, −9,576.8). (*B*) Interpopulation differentiation with pairwise F_{ST} estimates are above the diagonal and pairwise Φ_{ST} estimates are below the diagonal. All are significant at $P < 0.001$ except when indicated (*$P < 0.05$; ** $P < 0.01$).

Discussion

We investigated the evolutionary history of Bornean orangutans using the most comprehensive Bornean sample set compiled to date to our knowledge. Our mtDNA results indicate a surprisingly recent origin for current Bornean populations, and together with the nuclear markers, illustrate that their current distribution has been uniquely shaped by a combination of historical, geographical, and sociobehavioral factors.

Historical Factors: Recent Radiation of Bornean Populations. The recent coalescence of Bornean orangutan haplotypes in the Middle to Late Pleistocene is in striking contrast with that of the other Bornean canopy-bound rainforest species for which data are available, the gibbon *Hylobates muelleri.* This gibbon, distributed throughout the north, east, and west of Kalimantan, has a time to the most recent common ancestor (TMRCA) of 1.78 Ma (95% CI, 1.33–2.25) (36), suggesting that Bornean gibbons have been differentiating within the island for much longer than orangutans. Moreover, Sulawesi macaques (genus *Macaca*), whose ancestors dispersed from Borneo, have a TMRCA with their Bornean sister species of approximately 2 Ma (37). Although the exact timing of their migration is uncertain, the older mtDNA coalescence dates for both Bornean gibbons and Bornean and Sulawesi macaques suggests they have been in Borneo as far back as the Early Pleistocene. Therefore, it is conceivable that orangutans also arrived in Borneo around the same time. Yet, current Bornean orangutan mtDNA haplotypes stem from a very recent common ancestor originating in the Middle to Late Pleistocene.

The relatively short time to the most recent common ancestor of Bornean haplotypes is particularly striking given the deep Bornean–Sumatran orangutan coalescence approximately 3.5 Ma. Such a long differentiation between Bornean and Sumatran haplotypes appears hard to reconcile with the recent episodes of interconnectedness between the islands during the Pleistocene glaciations, most notably during the Last Glacial Maximum approximately 17 ka (2, 5). However, the presence of land bridges does not necessarily imply suitable conditions for migration. A savannah corridor (8) combined with riverine barriers dissecting the exposed land (5, 11) would have presented severe obstacles to migration for orangutans, restricting them to riverine forest galleries along the banks. Coalescence for Bornean and Sumatran haplotypes is expected to vary across species, reflecting differences in dispersal abilities, habitat requirements, or ancestral effective population size, aside from possible discrepancies in dating methods (38). For instance, the south Bornean gibbon *Hylobates albibarbis* and the Sumatran–Malaysian gibbon *Hylobates agilis* have a TMRCA of 1.56 Ma (36), and Bornean and Sumatran pigtailed macaques have one of 3 to 4 Ma (37). By contrast, the Bornean–Sumatran common ancestor of both the silvered langur

(39) and clouded leopard (40) is much more recent than that of orangutans, gibbons, and pig-tailed macaques, probably because of a higher flexibility in habitat use.

Assuming that orangutans arrived in Borneo around the same time as gibbons and macaques, the recent coalescence of Bornean orangutans could be explained by a bottleneck through a severe rainforest contraction. Such a bottleneck would have had a more dramatic impact on the mtDNA structure of orangutans compared with other species as a result of their low densities and slow life histories (18) as well as habitat requirements. Gibbons were apparently not affected by habitat changes as harshly, perhaps because populations can survive in smaller patches. Our findings are consistent with the survival and expansion of a single lineage from within a refugium in Borneo. Geomorphological and palynological data indicate the presence of dryer, more open vegetation in southern and western Borneo during the last glaciation (2, 41), and by extrapolation also during other glaciations (but c.f. refs. 42, 43). Climate change was especially severe during an extended cold period within the penultimate glaciation between 130 and 190 ka (44, 45), which occurred approximately at the time of mean coalescence of Bornean mtDNA haplotypes. More recently, the last Toba eruption approximately 74 ka resulted in a short, albeit significant, decrease in regional temperatures, ensued by a 1,800-y cold stadial (9, 10). Our data do not provide clear signals to make conclusive statements about potential Toba effects. Nonetheless, the coldest period of the penultimate glaciation (44, 45) was more prolonged than the cold period following the last Toba eruption, suggesting more severe effects of the former on the extent of rainforest across Sundaland. In any event, suitable rainforest habitat for orangutans should have existed in certain regions in Borneo where a refugium population survived the dry glacial conditions. Possible Pleistocene refugia in Borneo have also been described for numerous other rainforest species such as termites, ants, orchids, oaks, and large-bodied mammals (37, 46–51), and together with the isolation of the island, could act as a mechanism of evolutionary diversification driving high Bornean species endemism. Following the expansion of orangutans throughout the island, the Pleistocene climatic oscillations should have led to recurrent population expansions and contractions.

Geographical and Sociobehavioral Barriers. Despite the recent common ancestry of Bornean populations, our analyses revealed high and significant mitochondrial differentiation, with populations within currently recognized subspecies generally displaying as much differentiation as those between subspecies. Of notable interest is the great extent of subdivision and lack of reciprocal monophyly for the morphologically recognized subspecies *P. p. morio* and *P. p. wurmbii*. MtDNA haplotype sharing is uncommon and for populations separated by rivers occurs only in two instances: (i) for SA and GP and (ii) for the northern and southern populations across the Kinabatangan river. In both cases, very recent common ancestry could explain the incomplete mtDNA lineage sorting. For North Kinabatangan (NK) and SK, Jalil et al. (27) proposed an expansion from a recent common refugium further west in Mount Kinabalu, as posited for other Bornean species (46, 47, 49). DV, with its low haplotype diversity, might also be the result of a recent range expansion. GP is located proximally to the Bangka–Belitung–Karimata–Schwaner divide, from where orangutans are presumed to have dispersed to the rest of Borneo (12) and where we might expect a rich haplotype diversity. However, the presence of only one mtDNA haplotype shared with populations further east suggests that the current population in GP is recent and/or underwent a severe recent bottleneck. This and other local bottlenecks make it impossible to reconstruct a colonization of Borneo through the southwestern "choke point" (52).

The rarity of mtDNA haplotype sharing among Bornean populations contrasts with patterns in the patrilocal chimpanzees and bonobos (53, 54), where mtDNA sharing is extensive. Interestingly, two orangutan haplotypes from one site (Tuanan, TU) that were more closely related to those of another site (Sungai Lading, SL) pertain only to males. Although nuclear differentiation among the orangutan populations is significant, we find evidence for a small degree of nuclear gene flow, suggesting that it is male-mediated. Furthermore, the effect of rivers on the isolation by distance patterns for the mtDNA indicate that these are important barriers to female movement, probably as a result of smaller dispersal distances of females (18). An association between mtDNA genetic distance and distances around rivers has also been found in gorillas (20), and a role for differential dispersal distances between the sexes has been posited for western lowland gorillas (55). Our results are consistent with the pattern of female philopatry and male-biased dispersal proposed by Delgado and van Schaik (18) and indicate that the orangutan sexes are subject to very different constraints on mobility. Although female philopatric behavior may be responsible for the strong effect of geographical barriers on mtDNA structure, we cannot make any conclusive statements on the effects of rivers on males. More continuous sampling, especially along rivers and examination of Y-chromosomal markers, representative of male histories, will prove useful in determining how geographical barriers differentially affect the sexes. In addition, further geomorphological data on river course and width changes through time would contribute to the understanding of their vicariant action.

Bornean orangutan distribution and population structure has been uniquely shaped by the Pleistocene fluctuations and by sociobehavioral and geographical barriers to movement. Our findings support a recent radiation of Bornean orangutans in the Middle to Late Pleistocene, resulting in "static" clusters of females strongly separated by geographical barriers and subject to high differentiation, with more mobile males exerting a homogenizing influence on the nuclear gene pool. Further sampling will help establish whether there is a marker specific pattern of clusters versus clines resulting from sex-biased dispersal (c.f ref. 52). In addition, in depth population genetic studies of other endangered and endemic taxa such as the Bornean gibbons and Sumatran orangutans will be of interest in contrasting the differential effects of environmental processes.

Materials and Methods

Samples and Datasets. Our data comprise noninvasively collected fecal and hair samples from a number of long-term study sites: Gunung Palung (GP), Sabangau (SA), Sungai Lading (SL), Tuanan (TU), Danum Valley Conservation Area (DV), and the Lower Kinabatangan Wildlife Sanctuary (Fig. 1*B*). We partitioned the latter site into South Kinabatangan (SK) and North Kinabatangan (NK), given the significant differentiation between the locales found by Goossens et al. (28). In addition, we incorporated scattered samples from Warren et al. (17) (Table S2), encompassing most of the current distribution of *P. pygmaeus* (Fig. 1*B*). Depending on sample quality and data availability, we used two different datasets for mtDNA analyses, and two for nuclear microsatellite analyses (Table S3). DNA extraction and quantification procedures are described in *SI Materials and Methods*.

mtDNA Analyses. Based on unique microsatellite genotypes or mtDNA haplotypes (*SI Materials and Methods*), we obtained the following long-term study site sample sizes: SA (n = 23), SL (n = 26), TU (n = 30), and DV (n = 18). We also sequenced low DNA quantity samples from GP (n = 20), where individual identification was done through long-term observational data. Additionally, haplotypes for individuals from SK (n = 38) and NK (n = 35) were from Jalil et al. (27) (GenBank accession numbers EU547189–EU547201). Finally, we resequenced 21 extracts from the Bornean samples in Warren et al. (17) to cover the same region of mtDNA (Table S2). We sequenced a 323-bp region of the mtDNA hypervariable region I (HVRI). Details on the primers and PCR conditions and raw data analyses are provided in *SI Materials and Methods*. Summary statistics including haplotype diversity (h_d), nucleotide diversity (π), and average pairwise differences were calculated in DNAsp 5 (56) and Arlequin 3.11 (57). We conducted model selection tests on jModelTest 0.1 (58, 59), using the Akaike information criterion to choose the most suitable model and its parameters.

For the phylogenetic analyses, we incorporated HVRI haplotypes from all long-term study sites as well as Warren resequenced samples (Tables S2 and S3). First, to infer the coalescence date for Bornean mtDNA haplotypes, we used a Bayesian Markov chain Monte Carlo analysis as implemented in BEAST 1.5.4 (60) and produced a phylogenetic tree. We included the collapsed haplotypes from 211 Bornean and six Sumatran orangutans, as well as 19 humans as an outgroup. Based on the Akaike information criterion from jModeltest, we selected the HKY + G model. We used an uncorrelated relaxed log-normal clock (61), specifying a normal distribution with a mean HVRI substitution rate of 0.1643 substitutions per nucleotide per Myr for the mean rate prior. We chose this corrected HVRI estimate (62) because it takes into account the effects of purifying selection on the entire mtDNA molecule as well as saturation factors affecting the molecular rate decay described in numerous studies (38, 63, 64), and is therefore appropriate for population-level analyses (62, 65). The 95% confidence interval for the normal distribution spanned HVRI substitution rates obtained in other studies, from 0.06 to 0.25 substitutions/site/Myr (66). Using the birth-death prior for branching rates, we carried out two runs for 25 million generations with parameter sampling every 1,000 generations. Tracer 1.4.1 (67) was then used to examine whether the 10% burn-in period and effective sample sizes were adequate. Both runs were combined in LogCombiner 1.4.8, and the resulting tree visualized and edited using Figtree 1.2 (68), omitting human haplotypes. Second, to infer the coalescence date for Bornean and Sumatran mtDNA haplotypes, we used the same procedure, but instead of the corrected mutation rate, we chose two fossil based divergence estimates as priors. Fossil calibration points provide estimates of phylogenetic rates suitable for analyses at the interspecific level (65). The two calibration points were the Ponginae-Homininae divergence at approximately 14 Ma (69, 70) and the *Pan-Homo* divergence older than 6 Ma (71, 72). We specified log-normally distributed priors, appropriate for paleontological data (73). For the Ponginae-Homininae divergence, we used a log-normal mean of 0, log-normal SD of 0.56, and offset of 13 Ma, thereby obtaining a broad distribution with a 95% interval from 13.4 to 20 Ma. This range incorporates the uncertainties associated with the upper bound estimate of a split. For the *Pan-Homo* calibration, we used a log-normal mean of 0, log-normal SD of 0.56, and offset of 5 Ma, spanning a 95% interval from 5.4 to 7.5 Ma. The tree topology remained the same as in the first analysis, so it is not presented. Third, we investigated phylogenetic relationships at the intraspecific level by generating a median-joining network for the Bornean haplotypes using Network 4.0 (74).

For the population structure analyses, we used data from the long-term study sites GP, SA, SL, TU, DV, NK, and SK. In addition, we incorporated Danau Sentarum (DS) and Kutai (KU) sampling sites from Warren et al. (17) for which at least three samples of precise origin are available (cf. ref. 20; Table S2). We calculated pairwise Φ_{ST} values in Arlequin, using the Tamura Nei model (75) and a γ distribution shape parameter of 0.344. We obtained significance levels using 10,000 permutations. To define the most differentiated groups of populations, we also performed a spatial analysis of molecular variance (SAMOVA) with SAMOVA software, version 1.0 (34), using previously published geographical coordinates (17, 76).

Microsatellite Analyses. Microsatellite analyses focused only on samples from long-term study sites GP, SA, SL, TU, DV, SK, and NK. For the low DNA quality and quantity samples from GP, we could obtain genotypes for six individuals. We genotyped samples from all sites except SK and NK using a panel of 25 highly polymorphic nuclear microsatellite markers (28, 77) listed in Table S4, following the protocol given in *SI Materials and Methods*. Additionally, we incorporated previously generated data from NK and SK for 12 microsatellite markers (28), which were part of our panel of 25 markers. We standardized the data and performed identity analyses as described in *SI Materials and Methods*. After this procedure, we obtained two data sets: (*i*) dataset I includes 25 markers and 98 individuals from the five study sites GP (n = 6), SA (n = 19), SL (n = 26), TU (n = 29), and DV (n = 18); and (*ii*) dataset II includes 12 markers and 295 individuals from seven study sites, including all from dataset I plus NK (n = 91) and SK (n = 106).

After Bonferroni correction, we found no deviation from Hardy–Weinberg equilibrium, and only four pairs of different loci from two populations showed linkage disequilibrium, which is most likely explained by demographic effects rather than linkage. Also, we found evidence for possible null alleles for one locus in one population. As it was not consistent across populations, we did not exclude this locus from further analyses.

We used Genetix 4.05 (78) to obtain population pairwise F_{ST} values and significance levels. We also performed two separate analyses on Structure 2.3 (79) using the admixture model with correlated allele frequencies, and the Locprior model, which improves clustering when the signal is weak without spuriously inferring structure if absent (80). We specified a burn-in length of 10^5 followed by 10^6 Markov chain Monte Carlo steps. For each K, we ran the analysis 10 times. In the first analysis, we incorporated the widely distributed seven populations genotyped at 12 microsatellite markers (dataset II). In the second analysis, we further refined our findings focusing on the five populations for which we have genotypes for 25 microsatellite markers (dataset I).

We calculated geographical distance matrices as Euclidean and cost path distances between all study populations. The latter, representing true surface distances circumnavigating riverine barriers, were computed from the Shuttle Radar Topography Mission global Digital Elevation Model, as distributed by ESRI (81). We clipped the Digital Elevation Model to encompass the whole of Borneo and filled sinks to obtain a depressionless elevation model, which was then reprojected into the Universal Transverse Mercator coordinate system with a resolution of 100 m. From this, we constructed a flow accumulation raster and extracted grid cells with values of at least 1,000 to generate a stream order raster following the convention of Strahler (82). We then produced a cost raster by designating areas with flow accumulation values lower than 1,000 and streams of order 1 to 2, a cost of 1, whereas streams of orders 3, 4, and 5 were assigned costs of 3,000, 4,000, and 5,000, respectively. Streams of order 6 to 7 were designated as uncrossable barriers (cf. ref. 20). After masking the resulting cost raster with the Shuttle Radar Topography Mission Water dataset (81), we calculated dyadic cost path distances between the study populations. For all geospatial analyses, we used ArcInfo Spatial Analyst extension for ArcGIS 9.3 (83).

To investigate the association between genetic (pairwise Φ_{ST} for HVRI and F_{ST} for microsatellite markers) and geographical distances (Euclidean and cost path), we performed (partial) Mantel tests in R 2.10.1 (84), using the "ecodist" package (85).

ACKNOWLEDGMENTS. We thank the Indonesian Institute of Sciences, the Indonesian State Ministry for Research and Technology, and the Sabah Wildlife Department for granting permission to undertake this research. We are indebted to all individuals who helped collecting samples in the field. Samples were exported from Indonesia, Malaysia, and the U.K. to Switzerland under permits 07279/IV/SATS-LN/2009, 00961/IV/SATS-LN/2007, 4645, 06968/IV/SATS-LN/2005, and 290569/01 from the Convention on International Trade in Endangered Species of Wild Fauna and Flora. We thank C. Zollikofer for insightful comments. We are also grateful to S. Ho and A. Drummond for assistance with methods, M. Greminger for assistance with editing, and M. Gasan for assistance with graphics. Funding for this study was obtained from Swiss National Science Foundation Grant 31003A-116848 (to M.K. and C.P.v.S.), the Messerli Foundation, Claraz Schenkung, and A. H. Schultz-Stiftung.

1. Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philos Trans R Soc Lond B Biol Sci* 359:183–195.
2. Bird MI, Taylor D, Hunt C (2005) Palaeoenvironments of insular Southeast Asia during the last glacial period: A savanna corridor in Sundaland? *Quat Sci Rev* 24:2228–2242.
3. Hall R (2002) Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: Computer-based reconstructions, model and animations. *J Asian Earth Sci* 20:353–431.
4. Verstappen HT (1997) The effect of climatic change on southeast Asian geomorphology. *J Quaternary Sci* 12:413–418.
5. Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *J Biogeogr* 27:1153–1167.
6. Fordham D, Brook B (2010) Why tropical island endemics are acutely susceptible to global change. *Biodivers Conserv* 19:329–342.
7. Verstappen HT (1975) On paleo-climates and landform development in Malesia. *Modern Quaternary Res Southeast Asia* 1:3–35.
8. Heaney LR (1991) A synopsis of climatic and vegetational change in Southeast Asia. *Clim Change* 19:53–61.
9. Rampino MR, Self S (1992) Volcanic winter and accelerated glaciation following the Toba super-eruption. *Nature* 359:50–52.
10. Williams MAJ, et al. (2009) Environmental impact of the 73 ka Toba super-eruption in South Asia. *Palaeogeogr Palaeoclimatol Palaeoecol* 284:295–314.
11. Harrison T, Krigbaum J, Manser J (2006) Primate biogeography and ecology on the Sunda Shelf islands: A paleontological and zooarchaeological perspective. *Primate Biogeography*, eds Lehman SM, Fleagle JG (Springer, New York), pp 331–374.
12. Rijksen HD, Meijaard E (1999) *Our Vanishing Relative: The Status of Wild Orangutans at the Close of the Twentieth Century* (Kluwer, Dordrecht, The Netherlands).
13. Slik JWF, et al. (2003) A floristic analysis of the lowland dipterocarp forests of Borneo. *J Biogeogr* 30:1517–1531.
14. Moss SJ, Wilson MEJ (1998) Biogeographic implications from the Tertiary palaeogeographic evolution of Sulawesi and Borneo. *Biogeography and Geological Evolution of SE Asia*, eds Hall R, Holloway JD (Backhuys, Leiden, The Netherlands), pp 133–155.
15. MacKinnon K, Hatta G, Halim H, Mangalik A (1996) The island of Borneo. *The Ecology of Kalimantan. The Ecology of Indonesia Series* (Periplus Editions, Singapore), Vol 3, pp 9–68.

16. van Welzen PC, Slik JWF, Alahuhta J (2005) Plant distribution patterns and plate tectonics in Malesia. *Biol Skr* 55:199–217.
17. Warren KS, et al. (2001) Speciation and intrasubspecific variation of Bornean orangutans, *Pongo pygmaeus pygmaeus*. *Mol Biol Evol* 18:472–480.
18. Delgado RA, Van Schaik CP (2000) The behavioral ecology and conservation of the orangutan (Pongo pygmaeus): A tale of two islands. *Evol Anthropol* 9:201–218.
19. van Schaik CP, et al. (2003) Orangutan cultures and the evolution of material culture. *Science* 299:102–105.
20. Anthony NM, et al. (2007) The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proc Natl Acad Sci USA* 104:20432–20436.
21. Eriksson J, Hohmann G, Boesch C, Vigilant L (2004) Rivers influence the population genetic structure of bonobos (*Pan paniscus*). *Mol Ecol* 13:3425–3435.
22. Gonder MK, Disotell TR, Oates JF (2006) New genetic evidence on the evolution of chimpanzee populations and implications for taxonomy. *Int J Primatol* 27:1103–1127.
23. Xu X, Arnason U (1996) The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. *J Mol Evol* 43:431–437.
24. Muir CC, Galdikas BMF, Beckenbach AT (2000) mtDNA sequence diversity of orangutans from the islands of Borneo and Sumatra. *J Mol Evol* 51:471–480.
25. Steiper ME (2006) Population history, biogeography, and taxonomy of orangutans (Genus: *Pongo*) based on a population genetic meta-analysis of multiple loci. *J Hum Evol* 50:509–522.
26. Groves CP (2001) *Primate Taxonomy* (Smithsonian Institution Press, Washington, DC).
27. Jalil MF, et al. (2008) Riverine effects on mitochondrial structure of Bornean orangutans (*Pongo pygmaeus*) at two spatial scales. *Mol Ecol* 17:2898–2909.
28. Goossens B, et al. (2005) Patterns of genetic diversity and migration in increasingly fragmented and declining orangutan (*Pongo pygmaeus*) populations from Sabah, Malaysia. *Mol Ecol* 14:441–456.
29. Singleton I, et al. (2004) *Orangutan Population and Habitat Viability Assessment: Final Report* (IUCN/SSC Conservation Breeding Specialist Group, Apple Valley, MN).
30. Hofreiter M, Stewart J (2009) Ecological change, range fluctuations and population dynamics during the Pleistocene. *Curr Biol* 19:R584–R594.
31. Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. *Trends Ecol Evol* 23:564–571.
32. Posada D, Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
33. Posada D, Crandall KA (2001) Intraspecific gene genealogies: Trees grafting into networks. *Trends Ecol Evol* 16:37–45.
34. Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Mol Ecol* 11:2571–2581.
35. Jost L (2008) G(ST) and its relatives do not measure differentiation. *Mol Ecol* 17: 4015–4026.
36. Thinh VN, et al. (2010) Mitochondrial evidence for multiple radiations in the evolutionary history of small apes. *BMC Evol Biol* 10:74.
37. Ziegler T, et al. (2007) Molecular phylogeny and evolutionary history of Southeast Asian macaques forming the M. silenus group. *Mol Phylogenet Evol* 42:807–816.
38. Pulquério MJF, Nichols RA (2007) Dates from the molecular clock: How wrong can we be? *Trends Ecol Evol* 22:180–184.
39. Roos C, Nadler T, Walter L (2008) Mitochondrial phylogeny, taxonomy and biogeography of the silvered langur species group (Trachypithecus cristatus). *Mol Phylogenet Evol* 47:629–636.
40. Wilting A, et al. (2007) Clouded leopard phylogeny revisited: Support for species recognition and population division between Borneo and Sumatra. *Front Zool* 4:15.
41. Meijaard E (2004) *Solving mammalian riddles. A reconstruction of the Tertiary and Quaternary distribution of mammals and their palaeoenvironments in island South-East Asia.* (Australian National University, Canberra) PhD thesis.
42. Cannon CH, Morley RJ, Bush ABG (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proc Natl Acad Sci USA* 106:11188–11193.
43. Woodruff DS (2010) Biogeography and conservation in Southeast Asia: How 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. *Biodivers Conserv* 19:919–941.
44. Martinson DG, et al. (1987) Age dating and the orbital theory of the ice ages: Development of a high-resolution 0 to 300,000-year chronostratigraphy. *Quat Res* 27: 1–29.
45. Wright JD (2000) *Global Climate Change in Marine Stable Isotope Records. Quaternary Geochronology: Methods and Applications to Paleoseismology, AGU Reference Shelf Series*, eds Noller JS, Sowers JM, Lettis WR (American Geophysical Union, Washington, DC), pp 427–433.
46. Barkman TJ, Simpson BB (2001) Origin of high-elevation dendrochilum species (Orchidaceae) endemic to Mount Kinabalu, Sabah, Malaysia. *Syst Bot* 26:658–669.
47. Gathorne-Hardy FJ, Davies RG, Eggleton P, Jones DT (2002) Quaternary rainforest refugia in south-east Asia: Using termites (Isoptera) as indicators. *Biol J Linn Soc Lond* 75:453–466.
48. Quek SP, Davies SJ, Ashton PS, Itino T, Pierce NE (2007) The geography of diversification in mutualistic ants: A gene's-eye view into the Neogene history of Sundaland rain forests. *Mol Ecol* 16:2045–2062.
49. Cannon CH, Manos PS (2003) Phylogeography of the Southeast Asian stone oaks *(Lithocarpus). J Biogeogr* 30:211–226.
50. Meijaard E, Groves CP (2004) The biogeographical evolution and phylogeny of the genus *Presbytis*. *Primate Rep* 68:71–90.
51. Louys J (2007) Limited effect of the Quaternary's largest super-eruption (Toba) on land mammals from Southeast Asia. *Quat Sci Rev* 26:3108–3117.
52. Handley LJ, Manica A, Goudet J, Balloux F (2007) Going the distance: human population genetics in a clinal world. *Trends Genet* 23:432–439.
53. Langergraber KE, et al. (2007) The genetic signature of sex-biased migration in patrilocal chimpanzees and humans. *PLoS ONE* 2:e973.
54. Eriksson J, et al. (2006) Y-chromosome analysis confirms highly sex-biased dispersal and suggests a low male effective population size in bonobos (*Pan paniscus*). *Mol Ecol* 15:939–949.
55. Douadi MI, et al. (2007) Sex-biased dispersal in western lowland gorillas (*Gorilla gorilla gorilla*). *Mol Ecol* 16:2247–2259.
56. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
57. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
58. Posada D (2008) jModelTest: Phylogenetic model averaging. *Mol Biol Evol* 25: 1253–1256.
59. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
60. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
61. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88.
62. Soares P, et al. (2009) Correcting for purifying selection: An improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759.
63. Endicott P, Ho SYW, Metspalu M, Stringer C (2009) Evaluating the mitochondrial timescale of human evolution. *Trends Ecol Evol* 24:515–521.
64. Ho SY, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* 22:1561–1568.
65. Ho SY, Saarma U, Barnett R, Haile J, Shapiro B (2008) The effect of inappropriate calibration: Three case studies in molecular ecology. *PLoS ONE* 3:e1615.
66. Santos C, et al. (2005) Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). *Mol Biol Evol* 22:1490–1505.
67. Rambaut A, Drummond A (2005) *Tracer v1.3: MCMC Trace Analysis Tool* (Univ of Edinburgh, Edinburgh).
68. Rambaut A (2008) *FigTree: Tree Figure Drawing Tool, Version 1.2.2* (Univ of Edinburgh, Edinburgh).
69. Kelley J (2002) The Hominoid Radiation in Asia. *The Primate Fossil Record*, ed Hartwig W (Cambridge Univ Press, Cambridge), pp 369–384.
70. Raaum RL, Sterner KN, Noviello CM, Stewart C-B, Disotell TR (2005) Catarrhine primate divergence dates estimated from complete mitochondrial genomes: Concordance with fossil and nuclear DNA evidence. *J Hum Evol* 48:237–257.
71. Brunet M, et al. (2002) A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* 418:145–151.
72. Zollikofer CPE, et al. (2005) Virtual cranial reconstruction of Sahelanthropus tchadensis. *Nature* 434:755–759.
73. Ho SYM (2007) Calibrating molecular estimates of substitution rates and divergence times in birds. *J Avian Biol* 38:409–414.
74. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
75. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10: 512–526.
76. Wich S, Utami Atmoko S, Setia T, van Schaik C, eds (2009) *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (Oxford Univ Press, New York).
77. Nietlisbach P, Nater A, Greminger M, Arora N, Kruetzen M (2010) A multiplex-system to target 16 male-specific and 15 autosomal genetic markers for orangutans (genus: Pongo). *Conserv Genet Resources* 2:153–158.
78. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) *GENETIX 4.05, Logiciel sous Windows Pour la Génétique des Populations* (Université de Montpellier II, Montpellier, France), vol 4.05.
79. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data.. *Genetics* 155:945–959.
80. Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9: 1322–1332.
81. Environmental Systems Research Institute (2008) *ESRI Data & Maps 9.3* (Environmental Systems Research Institute, Redlands, CA).
82. Strahler AN (1957) Quantitative analysis of watershed geomorphology. *Trans Am Geophys Union* 8:913–920.
83. Environmental Systems Research Institute (2008) *ArcGis 9.3* (Environmental Systems Research Institute, Redlands, CA).
84. R Development Core Team (2009) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
85. Goslee SC, Urban DL (2007) The Ecodist package for dissimilarity-based analysis of ecological data. *J Stat Soft* 22:1–19.

Conservation Genet Resour (2010) 2:153–158
DOI 10.1007/s12686-010-9278-2

TECHNICAL NOTE

A multiplex-system to target 16 male-specific and 15 autosomal genetic markers for orang-utans (genus: *Pongo*)

Pirmin Nietlisbach · Alexander Nater · Maja P. Greminger · Natasha Arora · Michael Krützen

Received: 30 June 2010 / Accepted: 6 July 2010 / Published online: 25 July 2010

Abstract Genetic studies of dispersal on local spatial and short temporal scales require a large number of autosomal microsatellites. However, the study of dispersal over large spatial scales and the resolution of deep evolutionary histories require marker systems that are preferentially inherited through the male or female line. Addressing such questions in endangered orang-utans (genus: *Pongo*) bears significant relevance to species conservation, as habitat destruction and fragmentation pose a significant threat to the whole genus. Here, we report 16 male-specific markers (nine human-derived microsatellites, six single nucleotide and one insertion-deletion polymorphisms), and 15 novel *Pongo*-derived autosomal microsatellite loci. All 31 markers can be amplified in four multiplex polymerase chain reactions even in DNA derived from faecal material. The markers can be applied to studying a wide range of important questions in this genus, such as conservation genetics, social structure, phylogeny and phylogeography.

Keywords *Pongo* spp. · Single nucleotide polymorphisms · Microsatellites · Y chromosome · SNP typing · Non-invasive samples

Electronic supplementary material The online version of this article (doi:10.1007/s12686-010-9278-2) contains supplementary material, which is available to authorized users.

P. Nietlisbach · A. Nater · M. P. Greminger · N. Arora · M. Krützen (✉)
Evolutionary Genetics Group, Anthropological Institute and Museum, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
e-mail: michael.kruetzen@aim.uzh.ch

The endangered orang-utans occur on the islands of Borneo (*Pongo pygmaeus*; about 50,000 animals) and Sumatra (*P. abelii*; about 6,500 animals), where they have undergone a recent dramatic decline in population size (Goossens et al. 2006; Wich et al. 2008). This has been mostly attributed to habitat loss, leading to heavily fragmented populations of often only a few hundred individuals (Wich et al. 2008). Therefore, it is essential to maintain genetic diversity, which has been linked to population fitness (e.g. Reed and Frankham 2003). This can be achieved by maintaining corridors between fragmented populations, allowing animals to follow natural dispersal patterns (Gilbert-Norton et al. 2010).

Studying natural dispersal in wild orang-utans pose significant challenges. Behavioural observations suggested higher male than female dispersal (Delgado and van Schaik 2000), although this has not been fully confirmed by previous genetic studies (Utami et al. 2002; Goossens et al. 2005), where patterns of direct dispersal were investigated using autosomal microsatellite markers. However, direct inferences from autosomal markers are limited to the timescale of a few generations and geographically small areas, as sexual recombination will break down sex-specific information (Goudet et al. 2002). Sex-biased dispersal over larger time and spatial scales can be investigated by contrasting genetic information obtained from markers inherited through either the male or female lineage (Handley and Perrin 2007).

In orang-utans, maternally transmitted mitochondrial DNA markers are widely available (e.g. Warren et al. 2001), but markers on the male-specific region of the Y chromosome have not yet been applied. Here, we report 16 male-specific markers for the application in the genus *Pongo*. Nine of these markers are human-derived microsatellite loci, six are single nucleotide polymorphisms (SNPs) and one is an

insertion-deletion polymorphism (indel). We also provide 15 novel orang-utan specific autosomal microsatellite markers, which yielded more reliable amplification results for use with highly degraded faecal samples than human-derived markers (data not shown).

In a previous study (Erler et al. 2004), 45 male-specific human-derived loci were shown to amplify in orang-utans. Of those, we selected 25 loci re-sequenced them, designed orang-utan specific primers, and tested for levels of polymorphism in 14 Bornean and Sumatran orang-utans (Table 1; see laboratory procedures below). After aligning flanking regions of all 25 microsatellite loci, we discovered two SNPs and one indel. Additionally, we identified four more SNPs by sequencing Y chromosome conserved anchored tagged sequence loci DBY13, SMCY12, and SMCY14 (Hellborg and Ellegren 2003). We combined 16 Y-specific polymorphic markers (GenBank Acc.No. HM803995-HM804006) into two multiplex PCRs (Table 1). Actual repeat number was linked to amplicon length through sequence data (Table S1 in the Online Resources). For SNP typing, we applied a modified fragment length discrepant allele specific PCR (Li et al. 2009). Each PCR combined a fluorescently labelled forward primer with two interrogating reverse primers. The reverse primers were complementary to one of each SNP allele at their 3′-end and differed in their length at the 5′-end; the resulting difference in product length allowed for allelic discrimination by fragment length analysis (Fig. 1a). One microsatellite locus had a SNP in close proximity and another a SNP and an indel, which allowed for simultaneous screening in one or two PCRs, respectively (Fig. 1). Multiplex reactions for Y markers consisted of 1 µl DNA template in an 8 µl volume, and were carried out with the Qiagen PCR Multiplex Kit according to manufacturer's instructions and varying primer concentrations (Table 1). In addition to the manufacturer's recommendations, we also included a touchdown step in the PCRs, starting at 4°C above the reported annealing temperature in Table 1, then decreasing by 0.5°C per cycle for eight cycles to reach the annealing temperature for the last 40 cycles. We assessed the variability of the Y-markers on a population level in 21 and 14 orang-utans from one Bornean (subspecies *P. pygmaeus wurmbii*) and one Sumatran site, respectively. In addition, we report the number of Y-polymorphisms in an extensive sample of 173 animals covering most of the range of the genus *Pongo* (Table 1, Nietlisbach 2009). We tested scoring reliability by repeating 32% of all genotyping reactions, and found an average scoring error rate of 0.6% over all loci. It has not escaped our notice that our results immediately suggest a possible diagnostic system for elucidating species affiliation or hybrid status for males of this genus

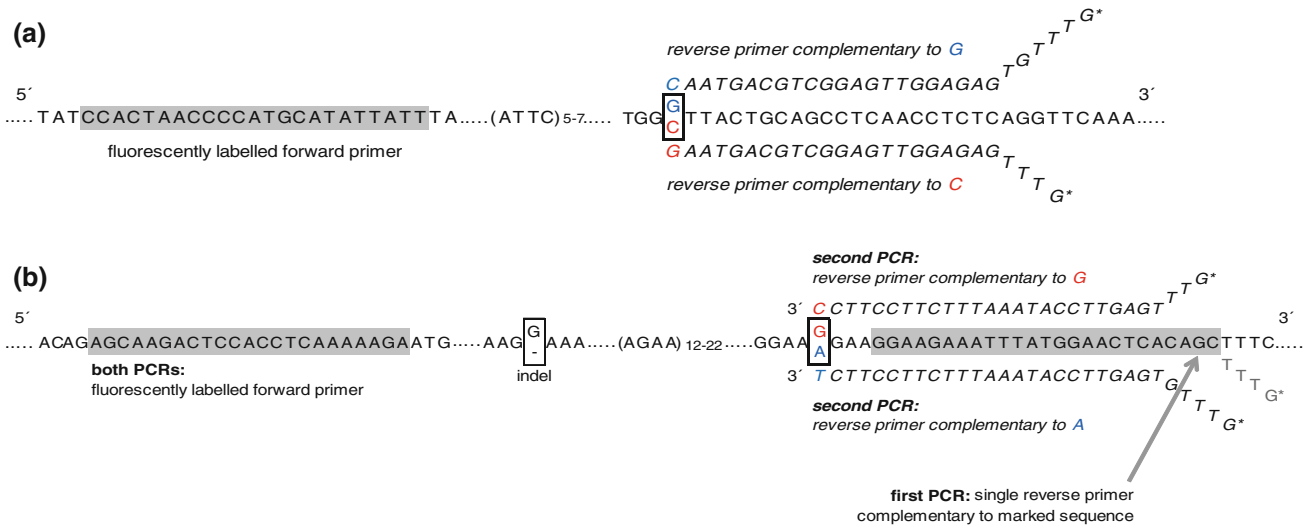


Fig. 1 Schematic representation of genotyping techniques for combined microsatellite/SNP typing. **a** Genotyping of SNP and microsatellite locus (DYS577). This locus was typed by a modified fragment length discrepant allele specific PCR (see text, Li et al. 2009) which targets the SNP in the box and the repeat motif of the microsatellite in one PCR. Discrimination between the two states of the SNP was achieved through competition of two reverse primers differing in length; and both reverse primers having a 1 to 2°C lower melting temperature and higher concentration than the fluorescently labelled forward primer (Table 1) to increase competition. Calibration was done with samples of known sequence. **b** Typing of three markers in two PCRs (DYS630). A first PCR using the grey primers amplified the region containing all three markers. A second PCR discriminates between SNP states as described in Fig. 1a. Due to confounding effects of the polymorphic microsatellite and indel, amplicon sizes of the first and second reactions have to be compared. If the amplicon of the first reaction is 5 bases longer than that of the second reaction, the SNP allele at this locus is G; if the difference is 3 bases, the SNP state is A. * PIG-tails were added to the 5′-end of reverse primers to enhance adenylation (Brownstein et al. 1996)

Table 1 Primers for male-specific and autosomal markers in orang-utans

Locus	Primer sequence 5′–3′	PC	T_A	Polymorphism	Overall		Suaq (Sumatra)				Tuanan (Borneo)			
					N	N_A	N	N_A	H_O	H_E	N	N_A	H_O	H_E
DYS630	*PET*-AGCAAGACTCCACCTCAAAAAGA*	0.15	63	AGAA	172	11	14	3		0.66	21	4		0.69
	gtttGCTGTGAGTTCCATAAATTTCTTCC	0.20		indel	172	2	14	1		0	21	1		0
DYS587	*6FAM*-AAAATTACCTTCTTTGGAAAGTAGTATT	0.30	63	ATACA	166	8	14	1		0	19	1		0
	gtttGTTATTTCTGAGCAGGGTTTCTAAG	0.40												
DYS532	*NED*-AGCAGGATTCCCTCTAAAAATATCA	0.10	63	compound,	171	3	14	1		0	21	2		0.09
	gTTTCTCCCTCCCTCCCTCTC	0.14		main motif (CTTT)										
DYS577	*6FAM*-CCACTAACCCCATGCATATTATT	0.30	63	GAAT	171	2	14	1		0	19	1		0
	gtttGAGAGGTTGAGGCTGCAGTAAG	0.40		C/G	171	2	14	2		0.13	19	1		0
	gtttgtGAGAGGTTGAGGCTG CAGTAAC	0.40												
DYS645	*6FAM*-GTACTAATTTTATTTCTTATGGCGTAGA	0.15	63	GTTTT	173	2	14	1		0	21	1		0
	gtttACACATGGCACCTGACACTG	0.20												
Y6C2	*6FAM*-CTTCTCCTTCTTCCTATTCTTCATCT	0.10	63	TTC	172	2	14	1		0	20	1		0
	gtttCAATAGTTTGGGGAAATAAGACAATG	0.14												
DBY13	*6FAM*-GGAAACTAAAATATGACATTGTAATTTG	0.30	63	C/G	168	2	14	1		0	20	1		0
	gtttATTTTTTTATTGTGATGCATACAGC	0.40												
	gtttgtATTTTTTTATTGTGATGCATACAGG	0.40												
DYS510	*PET*-GAAAGATAGATCAACAAGGTAGAAACAA	0.30	64	GATA	169	6	12	4		0.6	21	2		0.44
	gtttCATCCATCCATCCATCCATCT	0.40												
DYS561	*6FAM*-CCTGATGCCATCTGAAAATTAA	0.30	64	TAGA	168	5	14	1		0	20	3		0.52
	gtttACAACTGCACTCCAGCTTAGG	0.40												
DYS556	*6FAM*-TTACAAAACTAACATAAAGACCAAACAG	0.30	64	TAAA	172	3	14	1		0	21	2		0.41
	gtttGAAGCATTGAGTATAGTATAAAGTTGGT	0.40												
DYS630	*PET*-AGCAAGACTCCACCTCAAAAAGA*	0.15	64	A/G	171	2	14	1		0	21	1		0
	gttTGAGTTCCATAAATTTCTTCCTTCC	0.20												
	gtttgTGAGTTCCATAAATTTCTTCCTTCT	0.20												
SMCY12_26	*6FAM*-AAGGGTCACCACAGAAATACTTAG	0.15	64	C/G	173	2	14	2		0.13	21	1		0
	gtttgtCAGGTGGGGCGTAGTCTC	0.20												
	gtttCAGGTGGGGCGTAGTCTG	0.20												
SMCY12_337	*6FAM*-GTTACAGGTATACATGCACCTTTTT	0.15	64	A/C	171	2	14	1		0	21	1		0
	gtttgtGTGTGGCTTCTTTACTCTGTCA	0.20												
	gtttGTGTGGCTTCTTTACTCTGTCC	0.20												
SMCY14	*6FAM*-ATGGGGAAAAGATGAGTTCTGA	0.15	64	C/T	173	2	14	1		0	21	2		0.41
	gtttGTCTGGCATCCTAATGCCT	0.20												
	gtttgtGTCTGGCATCCTAATGCCC	0.20												

Table 1 continued

Locus	Primer sequence 5′–3′	PC	T_A	Polymorphism	Overall		Suaq (Sumatra)				Tuanan (Borneo)			
					N	N_A	N	N_A	H_O	H_E	N	N_A	H_O	H_E
O4_6	*PET*-GGCAATGTAACATATCCCTCTGTGT	0.05	58	GATA			23	4	0.61	0.68	28	3	0.71	0.62
	AGCCATGGACCTTGTGAGAAAAG	0.05												
O4_A5	*6FAM*-ATGGGCCCAGAAAACAACTCAGT	0.15	58	(GATA)(GATG)			22	4	0.64	0.55	26	6	0.65	0.65
	AGATAAAGGAATGGATAGATGGACAGA	0.15												
O4_A7	*VIC*-ACTGGCCCATTCAAAGTCTGTCATT	0.10	58	GTAG			21	4	0.86	0.72	26	2	0.35	0.29
	ACTGGCCCATTCAAAGTCTGT	0.10												
O4_A8	*NED*-CACAGGGTCCAAACTCAGATTATTG	0.20	58	(GATA)(GATG)			23	3	0.30	0.31	29	1	0	0
	CCTTCCCCCTCATGTAGTTATCAA	0.20												
O4_B5	*VIC*-GAGCCCTGATTCGTTTTACTGG	0.20	58	GATA			22	6	0.86	0.73	28	5	0.50	0.54
	AGCAAAGGCAGAAAACTGTAATGA	0.20												
O4_B6	*6FAM*-TGGAGCCTGAATATGTGACTGAAT	0.20	58	(GATA)(GTAG)			20	6	0.65	0.61	26	6	0.46	0.79
	AATGCCAGGATTTCCTTCTTTTT	0.20												
O4_B24	*6FAM*-TCTGAGGTACCCTGTAACAAAGAAA	0.10	58	GATA			23	3	0.65	0.56	29	1	0.00	0.00
	GAAATCCCAGTACCATATAAATGTCAT	0.10												
O4_A1	*6FAM*-CTCCCCTTCCTTCCTTTATTCAGTT	0.10	62	GTAG			23	5	0.87	0.73	28	4	0.82	0.75
	CAACACTTGGCAGTCACAAATCAG	0.10												
O4_B3	*VIC*-TTCCAGAAGGGGCGAGAAGTT	0.10	62	GACA			22	3	0.59	0.64	26	1	0	0
	GTTGGGACCAAACAGTTGTCAATAA	0.10												
O4_B17	*PET*-GTACCGACGGTGCACGAACAATGTA	0.30	62	GATG			19	3	0.68	0.67	26	6	0.69	0.73
	AGCCTGGCTGAAAAGTGGAACTGAG	0.30												
O4_B20	*NED*-CCTGCATTTTGTCACTCCCTCAACC	0.20	62	GATG			14	1	0	0	24	2	0.33	0.38
	CTGCCACACCTCCATGGACACAGAT	0.20												
O4_C9	*6FAM*-TGCAGGCCAGGGCTTCTTTCAA	0.15	62	GATA			22	5	0.55	0.54	27	4	0.59	0.63
	CAGTCTCCCCAGGACCCCTACACAG	0.15												
O4_C13	*6FAM*-CTGGGCACACTGTATATGGGGTAG	0.20	62	GATA			20	3	0.75	0.56	21	4	0.38	0.59
	GTTTGAGACCACTCATGATGCAAAGACC	0.20												
O4_Chr5	*PET*-CAGCAGCTCCTGAAATATCTGTCC	0.15	62	GATA			21	4	0.81	0.70	23	5	0.74	0.74
	GTTTGGGGTAGAGGAAAGCAGGTTGAT	0.15												
O4_Chr7	*NED*-CATCTCTTTATGGCTGACTGTTGAT	0.10	62	GATA			17	11	0.76	0.83	24	15	0.92	0.91
	GTTTGGTCCAAGACAAATTTGTATGAGT	0.10												

All loci with a Y in the name are Y-linked, all others are autosomal. Summary statistics are given for two study sites and over all sampled orang-utans. For loci DYS630 and DYS577, three and two male-specific markers were typed, respectively (Fig. 1). Loci combined in a single multiplex reaction have the same annealing temperature

Fluorescent labels are shown in italics at the 5′ end of the forward primer. PIG-tail bases (Brownstein et al. 1996) are given in lower case

PC primer concentration [μM], T_A annealing temperature, *N* number of samples, N_A number of different alleles, H_O observed heterozygosity, H_E expected heterozygosity (Nei 1987). * primer used in two PCRs (Fig. 1b)

Amplicon sizes and their relation to repeat numbers are shown in Table S1 in the Online Resources

(Nietlisbach 2009), if used in unison with readily available mtDNA markers.

To clone autosomal microsatellite markers, we extracted genomic DNA from 25 mg of frozen muscle tissue from a Sumatran orang-utan, using the DNeasy Tissue Kit (Qiagen). We digested ten micrograms of the purified DNA with *Nhe*I and *Alu*I (New England Biolabs) and size-selected for fragments between 400 and 1,200 base pairs length. Enrichment, cloning and sequencing were carried out as described in Nater et al. (2008), using only tetranucleotide biotinylated probes [$(GACA)_7$, $(GATA)_7$, and $(GATC)_7$]. We sequenced plasmids from 68 positive clones, of which 70% contained a microsatellite repeat. For 25 loci, which contained long uninterrupted repeats, we designed primers and amplified these loci in twelve orang-utans. Levels of polymorphism were qualitatively assessed on high-resolution Spreadex gels (Elchrom Scientific). Based on these results, we fluorescently labelled the forward primers of the 15 most polymorphic markers and combined these 15 loci (GenBank Acc.No. HM804007-HM804021) into two multiplex PCRs (Table 1). Then, we genotyped 29 orang-utans from Borneo and 23 from Sumatra, using DNA extracts from faecal samples with target DNA concentration ranging from 25 to 1,000 pg/µl, strictly following guidelines from Morin et al. (2001). PCRs using the Qiagen PCR Multiplex Kit contained 1 µl template DNA in an 8 µl final volume, with varying primer concentrations and annealing temperatures (Table 1). PCRs included 45 cycles with conditions according to manufacturer's instructions.

If not indicated otherwise, we used standard laboratory techniques at each step. We designed PCR primers with the PrimerSelect software implemented in Lasergene v7 (DNASTAR). PCR amplifications were performed on Veriti 96-well thermal cyclers (Applied Biosystems). Sequencing reactions were carried out using the BigDye Terminator v3.1 on a 3730 DNA Analyzer (both Applied Biosystems) according to manufacturer's instructions, cleaned-up using a $MgSO_4$ precipitation procedure, followed by resuspending the pellet in 20 µl ddH_2O. For fragment length analysis, PCR products were diluted 20–80 times in ddH_2O. One microlitre of this was added to 9.93 µl HiDi formamide and 0.07 µl of GeneScan 500 LIZ Size Standard (both Applied Biosystems) and denatured for three minutes at 95°C. We ran the samples on a 3730 DNA Analyzer and obtained genotypes using GeneMapper software v4.0 (Applied Biosystems). For the statistical analyses, we used MStools v3.1 add-into Microsoft Excel (Park 2001) and Genepop v4.0 (Rousset 2008).

Fragment length discrepant allele specific PCR used as SNP typing technique proved to be a reliable and cost-efficient strategy to assess SNP variation. The possibility to combine this technique with conventional microsatellite fragment length analysis makes it a suitable method to include a small number of SNPs to complement an extensive microsatellite analysis. The polymorphic male-specific markers for orang-utans described here promise to be highly useful for population genetic and phylogenetic studies addressing questions about dispersal strategies, phylogeographic patterns, and comparisons with other molecular markers. The autosomal markers can be applied to investigate local dispersal or assess relatedness and paternity. Knowledge about such processes, in particular about natural dispersal strategies, is important for species conservation.

Acknowledgments We thank C. van Schaik, M. van Noordwijk, J. Pamungkas, and D. Perwitasari-Farajallah. We are also indebted to all individuals who helped collecting samples in the field. This study was funded by the Swiss National Science Foundation (31003A-116848 to MK), Messerli Foundation, A.H.-Schultz Stiftung, and Claraz Schenkung. We thank the Indonesian Institute of Sciences (LIPI), the Indonesian State Ministry for Research and Technology (RISTEK), and the Sabah Wildlife Department for granting permission to undertake this research. All sampling and transportation of samples was conducted in accordance with Indonesian, Malaysian and international regulations (CITES).

References

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. Biotechniques 20(6): 1004–1010

Delgado RA, van Schaik CP (2000) The behavioral ecology and conservation of the orang-utan (*Pongo pygmaeus*): a tale of two islands. Evol Anthropol 9(5):201–218

Erler A, Stoneking M, Kayser M (2004) Development of Y-chromosomal microsatellite markers for nonhuman primates. Mol Ecol 13(10):2921–2930. doi:10.1111/j.1365-294X.2004.02304.x

Gilbert-Norton L, Wilson R, Stevens JR, Beard KH (2010) A meta-analytic review of corridor effectiveness. Conserv Biol 24(3): 660–668. doi:10.1111/j.1523-1739.2010.01450.x

Goossens B, Chikhi L, Jalil MF, Ancrenaz M, Lackman-Ancrenaz I, Mohamed M, Andau P, Bruford MW (2005) Patterns of genetic diversity and migration in increasingly fragmented and declining orang-utan (*Pongo pygmaeus*) populations from Sabah, Malaysia. Mol Ecol 14(2):441–456. doi:10.1111/j.1365-294X.02421.x

Goossens B, Chikhi L, Ancrenaz M, Lackman-Ancrenaz I, Andau P, Bruford MW (2006) Genetic signature of anthropogenic population collapse in orang-utans. PLoS Biol 4(2):285–291. doi: e2510.1371/journal.pbio.0040025

Goudet J, Perrin N, Waser P (2002) Tests for sex-biased dispersal using bi-parentally inherited genetic markers. Mol Ecol 11(6):1103–1114

Handley LJL, Perrin N (2007) Advances in our understanding of mammalian sex-biased dispersal. Mol Ecol 16(8):1559–1578. doi:10.1111/j.1365-294X.2006.03152.x

Hellborg L, Ellegren H (2003) Y chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male-specific DNA. Mol Ecol 12(1):283–291

Li SZ, Wan HR, Ji HY, Zhou KY, Yang G (2009) SNP discovery based on CATS and genotyping in the finless porpoise

(*Neophocaena phocaenoides*). Conserv Genet 10(6):2013–2019. doi:10.1007/s10592-009-9882-4

Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Mol Ecol 10(7):1835–1844

Nater A, Krützen M, Lindholm AK (2008) Development of polymorphic microsatellite markers for the livebearing fish *Poecilia parac*. Mol Ecol Resour 8(4):857–860. doi:10.1111/j.1755-0998.2008.02090.x

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nietlisbach P (2009) Male-specific markers in orang-utans (*Pongo* spp.)—Dispersal and phylogeny. MSc thesis. University of Zurich

Park SDE (2001) Trypanotolerance in West African cattle and the population genetic effects of selection. University of Dublin, Dublin, Ireland

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. Conserv Biol 17(1):230–237

Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8(1):103–106. doi:10.1111/j.1471-8286.2007.01931.x

Utami SS, Goossens B, Bruford MW, de Ruiter JR, van Hooff J (2002) Male bimaturism and reproductive success in Sumatran orang-utans. Behav Ecol 13(5):643–652

Warren KS, Verschoor EJ, Langenhuijzen S, Heriyanto, Swan RA, Vigilant L, Heeney JL (2001) Speciation and intrasubspecific variation of Bornean orang-utans, *Pongo pygmaeus pygmaeus*. Mol Biol Evol 18 (4):472-480

Wich SA, Meijaard E, Marshall AJ, Husson S, Ancrenaz M, Lacy RC, van Schaik CP, Sugardjito J, Simorangkir T, Traylor-Holzer K, Doughty M, Supriatna J, Dennis R, Gumal M, Knott CD, Singleton I (2008) Distribution and conservation status of the orang-utan (*Pongo* spp.) on Borneo and Sumatra: how many remain? Oryx 42(3):329–339. doi:10.1017/s003060530800197x

Curriculum Vitae

Name:	Alexander Nater
Date of birth:	November 10, 1981
Birthplace:	Zurich, Switzerland

Education

Aug. 2007 – June 2012	PhD in Evolutionary Biology, University of Zurich, Switzerland Dissertation: Processes Underlying Genetic Differentiation and Speciation in Orangutans (*Pongo* spp.), submitted March 2012
Feb. 2006 – June 2007	MSc in Anthropology, University of Zurich, Switzerland Thesis title: New microsatellite markers in orangutans (*Pongo* spp.) and bottlenose dolphins (*Tursiops* spp.)
Oct. 2001 – Nov. 2005	BSc in Biology, University of Zurich, Switzerland

Awards

- Award for outstanding scientific achievement from the Faculty of Science of the University of Zurich for the PhD thesis «Processes Underlying Genetic Differentiation and Speciation in Orangutans (*Pongo* spp.)», October 2012.

Selected Publications

- **Nater A**, Arora N, Greminger MP, van Schaik CP, Singleton I, Wich SA, Fredriksson G, Perwitasari-Farajallah D, Pamungkas J, Krützen M (2013). Marked Population Structure and Recent Migration in the Critically Endangered Sumatran Orangutan (*Pongo abelii*). Journal of Heredity, 104 (1): 2-13.
- Nietlisbach P, Arora N, **Nater A**, Goossens B, van Schaik CP, Krützen M (2012). Heavily male-biased long distance dispersal of orang-utans (genus: *Pongo*), as revealed by Y chromosomal and mitochondrial genetic markers. Molecular Ecology, 21: 3173-3186.
- **Nater A**, Nietlisbach P, Arora N, van Schaik CP, van Noordwijk MA, Willems EP, Singleton I, Wich SA, Goossens B, Warren KS, Verschoor EJ, Perwitasari-Farajallah D, Pamungkas J, Krützen M (2011). Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant orangutans (genus: *Pongo*). Molecular Biology and Evolution, 28(8): 2275-2288.
- Arora N, **Nater A**, van Schaik CP, Willems EP, van Noordwijk MA, Morf N, Krützen M, et al. (2010). Effects of Pleistocene glaciations and rivers on the population structure of Bornean orangutans (*Pongo pygmaeus*). Proceedings of the National Academy of Sciences of the United States of America (PNAS), 107(50): 21376-21381.